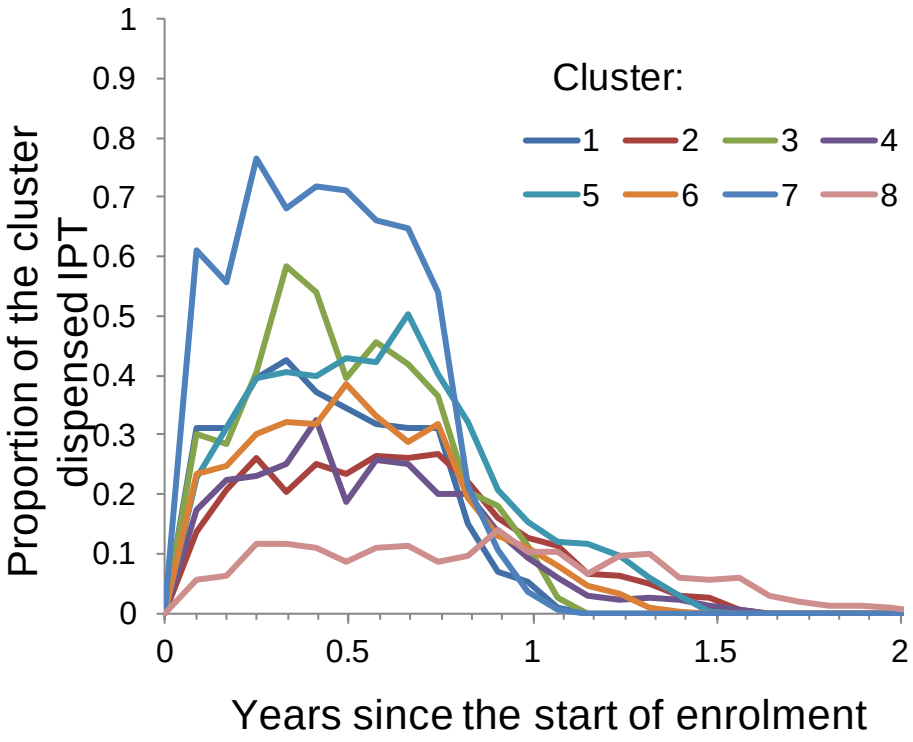
Tuberculosis control in South African goldmines: mathematical modelling of a trial of community-wide isoniazid preventive therapy

Web Appendix

Table of Contents



Web Figure 1: Reported proportion of miners in each intervention cluster that was on IPT by time since the start of the intervention.

Web Appendix 1: Model equations and definitions

1.1 Overview

The model was set up using weekly time steps using the difference equations below. The model was written using the C programming language. Web Table 1 summarizes the definitions of the compartments and variables in the model; Table 1 in the main text provides the main parameters and variables; any additional parameters are defined below.

The population was allowed to experience the benefits of IPT (i.e. reduced rates of disease onset) or lack of benefit in the same week as they started or stopped IPT respectively. To simplify the equations whilst allowing this to occur, the population in the IPT-related compartments was transferred into subsequent strata at the end of each time step, once other transitions had been accounted for.

1.2 Equations

1.2.1 Latent and reinfected compartments

To ensure that that no one in the population could start IPT multiple times, the latent and reinfected compartments are subdivided according to whether or not they have been on IPT previously. For simplicity, this detail is omitted from the model diagram (Figure 2 in the main text). However, the disease-related compartments have not been stratified according to previous IPT – this simplification is unlikely to affect conclusions since the long duration (6 months) for tuberculosis treatment means that a negligible proportion of the model population could experience IPT twice and treatment for tuberculosis disease.

Miners with Latent infection

$$L_{z-,a,h}(t+\delta t) = L_{z-,a,h}(t)(1-i_{z+,a}(t))(1-d_{n,a,h,z-}k_{si,a,h}-\lambda(t)-m_{tr-,a}) + M_{in,L,a,h}(t) + R_{z-,a,h}(t,T_R) \quad \text{Equation 1.1a}$$

$$L_{z+,a,h}(t+\delta t,0) = L_{z-,a,h}(t)i_{z+,a}(t)(1-d_{n,a,h,z+}k_{si,a,h}-(1-\pi_{r,z+,h})\lambda(t)-m_{tr-,a}) \quad \text{Equation 1.1b}$$

$$L_{z+,a,h}(t+\delta t, s_z+\delta s_z) = L_{z+,a,h}(t,s_z) - L_{z+,a,h}(t,s_z)(d_{n,a,h,z+}k_{si,a,h} + (1-\pi_{r,z+,h})\lambda(t) + m_{tr-,a}) + R_{z+,a,h}(t,T_R,s_z)$$ Equation 1.1c

Miners who have completed IPT but have not been reinfected in the previous 2 years

$$P_{e+,a,h}(t+\delta t) = P_{e+,a,h}(t) - (\lambda(t)+m_{tr-,a})P_{e+,a,h}(t) + V_{z+,a,h}(t,T_{z_{\max}})$$ Equation 1.2a

$$P_{e-,a,h}(t+\delta t) = P_{e-,a,h}(t) - P_{e-,a,h}(t)(d_{n,z-,a,h}k_{si,a,h} + \lambda(t) + m_{tr-,a}) + R_{z_p}(t,T_R) + L_{z+,a,h}(t,T_{z_{\max}})$$ Equation 1.2b

Reinfected miners who are not on IPT

$$R_{z-,a,h}(t+\delta t,0) = (1-i_{z+}(t)\lambda(t)(L_{z-,a,h}(t)+V_{z-,a,h}(t))(1-d_{x,z-,a,h}(0)k_{si,a,h})$$ Equation 1.3a

$$R_{z-,a,h}(t+\delta t, s_r+\delta s_r) = (1-i_{z+}(t))R_{z-,a,h}(t,s_r)(1-d_{x,z-,a,h}(s_r)k_{si,a,h} - m_{tr-,a}) + M_{in,R,a,h}(t,s_r)$$ Equation 1.3b

Reinfected miners who are on IPT

$$R_{z+,a,h}(t+\delta t, s_r+\delta s_r, 0) = R_{z-,a,h}(t,s_r)i_{z+,a}(t)(1-d_{x,z+,a,h,}(s_r)k_{si,a,h} - m_{tr-,a})$$ Equation 1.4a

$$R_{z+,a,h}(t+\delta t, 0, s_z) = (1-\pi_{r,z+,h})\lambda(t)(L_{z+,a,h}(t,s_z) + V_{z+,a,h}(t,s_z))(1-d_{x,z+,a,h,}(0)k_{si,a,h})$$ Equation 1.4b

$$R_{z+,a,h}(t+\delta t, s_r+\delta s_r, s_z+\delta s_z) = R_{z+,a,h}(t,s_r,s_z) - R_{z+,a,h}(t,s_r,s_z)(d_{x,z+,a,h,}(s_r)k_{si,a,h} + m_{tr-,a})$$ Equation 1.4c

Reinfected miners who have previously been on IPT

$$R_{z_p,a,h}(t+\delta t,0) = \lambda(t)(P_{e+,a,h}(t)+P_{e-,a,h}(t))(1-d_{x,z-,a,h}(0)k_{si,a,h})$$ Equation 1.5a

$$R_{z_p,a,h}(t+\delta t, s_r+\delta s_r) = R_{z_p,a,h}(t,s_r) - R_{z_p,a,h}(t,s_r)(d_{x,z-,a,h}(s_r)k_{si,a,h} + m_{tr-,a})$$ Equation 1.5b

1.2.2 Cases who have not yet been detected

To allow calculation of the proportion of tuberculosis cases that have been reinfected recently, cases which have not yet been detected are further stratified according to the mechanism by which they are experiencing disease (i.e. (exogenous) reinfection or (endogenous) reactivation). Once detected ("found"), cases remain in the detected compartments for a maximum period of 6 months (denoted by $T_{f_{\max}}$), unless they start treatment in the meantime, after which they are redistributed into the undetected compartments, according to their relative size. Considering cases experiencing disease

through endogenous reactivation, this is calculated using the equation $p_{En,s,a,h}(t) = \frac{E_{n,s,a,h}(t,T_{o_{\max}})}{E_{n,s,a,h}(t,T_{o_{\max}}) + E_{x,s,a,h}(t,T_{o_{\max}})}$. The equation considering cases of exogenous disease is analogous.

Cases experiencing disease because of reactivation, who have not yet been detected:

$$E_{n,s-,a,h}(t+\delta t,0) = L_{z-,a,h}(t) d_{n,z-,a,h} k_{si,a,h} + L_{z+,a,h}(t) d_{n,z+,a,h} k_{si,a,h} \qquad \text{Equation 1.6a}$$

$$\begin{aligned} E_{n,s-,a,h}(t+\delta t, s_o+\delta s_o) &= E_{n,s-,a,h}(t,s_o) \\ &- E_{n,s-,a,h}(t,s_o)(r_{f,s-,h,y} + r_{f,s-,h,p} + m_{tr-,a} + \mu_{s-,h,tr-} + o_{s+,h} + q(t)) \\ &+ (1-p_{in,f,s-}) M_{in,E_n,s-,a,h}(s_o) \end{aligned} \qquad \text{Equation 1.6b}$$

$$s_o < T_{o_{\max}}$$

$$\begin{aligned} E_{n,s-,a,h}(t+\delta t, s_o+\delta s_o) &= E_{n,s-,a,h}(t,s_o) \\ &- E_{n,s-,a,h}(t,s_o)(r_{f,s-,h,y} + r_{f,s-,h,p} + m_{tr-,a} + \mu_{s-,h,tr-} + o_{s+,h} + q(t)) \\ &+ (1-p_{in,f,s-}) M_{in,E_n,s-,a,h}(s_o) + p_{E_n,s-,a,h}(t) F_{s-,a,h}(t,T_{f_{\max}}) \end{aligned} \qquad \text{Equation 1.6c}$$

$$s_o = T_{o_{\max}}$$

$$\begin{aligned} E_{n,s+,a,h}(t+\delta t, s_o+\delta s_o) &= E_{n,s+,a,h}(t,s_o) \\ &- (r_{f,s+,h,y} + r_{f,s+,h,p} + m_{tr-,a} + \mu_{s+,h,tr-} + q(t)) E_{n,s+,a,h}(t,s_o) \\ &+ o_{s+,h} E_{n,s-,a,h}(t,s_o) + (1-p_{in,f,s+}) M_{in,E_n,s+,a,h}(t,s_o) \end{aligned} \qquad \text{Equation 1.6d}$$

$$s_o < T_{o_{\max}}$$

$$\begin{aligned} E_{n,s+,a,h}(t+\delta t, s_o+\delta s_o) &= E_{n,s+,a,h}(t,s_o) \\ &- (r_{f,s+,h,y} + r_{f,s+,h,p} + m_{tr-,a} + \mu_{s+,h,tr-} + q(t)) E_{n,s+,a,h}(t,s_o) \\ &+ o_{s+,h} E_{n,s-,a,h}(t,s_o) + (1-p_{in,f,s+}) M_{in,E_n,s+,a,h}(t,s_o) \\ &+ p_{E_n,s+,a,h}(t) F_{s+,a,h}(t,T_{f_{\max}}) \end{aligned} \qquad \text{Equation 1.6e}$$

$$s_o = T_{o_{\max}}$$

Cases experiencing disease because of reinfection, who have not yet been detected:

$$\begin{aligned} E_{x,s-,a,h}(t+\delta t,0) &= \sum_{s_r=0}^{T_R} d_{x,z-,a,h}(s_r) k_{si,a,h} (R_{z-,a,h}(t,s_r) + R_{z_p,a,h}(t,s_r)) \\ &+ \sum_{s_r=0}^{T_R} \sum_{s_z=0}^{T_{Z_{\max}}} R_{z+,a,h}(t,s_r,s_z) d_{x,z+,a,h}(s_r) k_{si,a,h} \end{aligned} \qquad \text{Equation 1.7a}$$

$$E_{x,s-,a,h}(t+\delta t, s_o+\delta s_o) = E_{x,s-,a,h}(t,s_o) - (r_{f,s-,h,y} + r_{f,s-,h,p} + m_{tr-,a} + \mu_{s-,h,tr-} + o_{s+,h} + q(t))E_{x,s-,a,h}(t,s_o) + (1-p_{in,f,s-})M_{in,E_x,s-,a,h}(t,s_o) \quad s_o < T_{o_{\max}}$$ Equation 1.7b

$$E_{x,s-,a,h}(t+\delta t, s_o+\delta s_o) = E_{x,s-,a,h}(t,s_o) - (r_{f,s-,h,y} + r_{f,s-,h,p} + m_{tr-,a} + \mu_{s-,h,tr-} + o_{s+,h} + q(t))E_{x,s-,a,h}(t,s_o) + (1-p_{in,f,s-})M_{in,E_x,s-,a,h}(t,s_o) + p_{E_x,s-,a,h}(t)F_{s-,a,h}(t,T_{f_{\max}}) \quad s_o = T_{o_{\max}}$$ Equation 1.7b

$$E_{x,s+,a,h}(t+\delta t, s_o+\delta s_o) = E_{x,s+,a,h}(t,s_o) - (r_{f,s+,h,y} + r_{f,s+,h,p} + m_{tr-,a} + \mu_{s+,h,tr-} + q(t))E_{x,s+,a,h}(t,s_o) + o_{s+,h}E_{x,s-,a,h}(t,s_o) + (1-p_{in,f,s+})M_{in,E_x,s+,a,h}(t,s_o) \quad s_o < T_{o_{\max}}$$ Equation 1.7c

$$E_{x,s+,a,h}(t+\delta t, s_o+\delta s_o) = E_{x,s+,a,h}(t,s_o) - (r_{f,s+,h,y} + r_{f,s+,h,p} + m_{tr-,a} + \mu_{s+,h,tr-} + q(t))E_{x,s+,a,h}(t,s_o) + o_{s+,h}E_{x,s-,a,h}(t,s_o) + (1-p_{in,f,s+})M_{in,E_x,s+,a,h}(t,s_o) + p_{E_x,s+,a,h}(t)F_{s+,a,h}(t,T_{f_{\max}}) \quad s_o = T_{o_{\max}}$$ Equation 1.7c

1.2.2.1 Detected cases:

$$F_{s-,a,h}(t+\delta t,0) = \sum_{s_o=0}^{T_{o_{\max}}}(r_{f,s-,h,y} + r_{f,s-,h,p} + q(t))(E_{n,s-a,h}(t,s_o) + E_{x,s-,a,h}(t,s_o)) + p_{in,f,s-}\sum_{s_o=0}^{T_{o_{\max}}}(M_{in,E_x,s-,a,h}(t,s_o) + M_{in,E_x,s-,a,h}(t,s_o))$$ Equation 1.8a

$$F_{s-,a,h}(t+\delta t, s_f+\delta s_f) = F_{s-,a,h}(t,s_f) - (m_{tr-,a} + \mu_{s-,h,tr-} + o_{s+,h} + \tau_{s-}(s_f))F_{s-,a,h}(t,s_f)$$ Equation 1.8b

$$F_{s+,a,h}(t+\delta t,0) = \sum_{s_o=0}^{T_{o_{\max}}}(r_{f,s+,y} + r_{f,s+,h,p} + q(t))(E_{n,s+,a,h}(t,s_o) + E_{x,s+,a,h}(t,s_o)) + p_{in,f,s+}\sum_{s_o=0}^{T_{o_{\max}}}(M_{in,E_x,s+,a,h}(t,s_o) + M_{in,E_x,s+,a,h}(t,s_o))$$ Equation 1.8c

$$F_{s+,a,h}(t+\delta t, s_f+\delta s_f) = F_{s+,a,h}(t,s_f) - F_{s+,a,h}(t,s_f)(m_{tr-,a} + \mu_{s+,h,tr-} + \tau_{s+}(s_f)) + o_{s+,h}F_{s-,a,h}(t,s_f) \quad 0 < s_f < T_{f_{\max}}$$ Equation 1.8d

1.2.2.2 Cases undergoing TB treatment:

$$C_{a,h}(t+\delta t,0)=\sum_{s_f=0}^{T_{f\max}}(\tau_{s-}(s_f)F_{s-,a,h}(t,s_f)+\tau_{s+}(s_f)F_{s+,a,h}(t,s_f)$$ Equation 1.9a

$$C_{a,h}(t+\delta t,s_\tau+\delta s_\tau)=C_{a,h}(t,s_\tau)-C_{a,h}(t,s_\tau)(m_{tr+}+\mu_{h,tr+})$$ Equation 1.9b

$$0<s_\tau<T_{\tau_{\max}}$$

1.2.2.3 Miners who have recovered from TB disease:

$$V_{z-,a,h}(t+\delta t)=V_{z-,a,h}(t)-V_{z-,a,h}(t)(i_{z+,a}(t)+\lambda(t)+m_{tr-,a})$$ Equation 1.10a
$$+M_{in,V,a,h}(t)+C(T_{\tau_{\max}})$$

$$V_{z+,a,h}(t+\delta t,0)=V_{z-,a,h}(t,s_z)i_{z+,a}(t)$$ Equation 1.10b

$$V_{z+,a,h}(t+\delta t,s_z+\delta s_z)=V_{z+,a,h}(t,s_z)$$ Equation 1.10c
$$-V_{z+,a,h}(t,s_z)(i_{z-,a}(t)+\lambda(t)(1-\pi_{r,z+,h})+m_{tr-,a})$$
$$0<s_z<T_{z_{\max}}$$

1.2.2.4 Transitions at the end of each time step:

$$P_{e+,a,h}(t+\delta t)=P_{e+,a,h}(t)+R_{z+,a,h}(t,T_{z_{\min}},T_{z_{\max}})p_{z_c,h}+$$ Equation 1.11a
$$+L_{z+,a,h}(t,T_{z_{\min}})i_{z-}(T_{z_{\min}})p_{z_c,h}+\sum_{s_z=0}^{T_{z_{\max}}}V_{z+,a,h}(t,s_z)i_{z-}(s_z)$$
$$+\sum_{s_r\geq T_{z_{\min}}}R_{z+,a,h}(t,s_r,T_{z_{\min}})i_{z-}(T_{z_{\min}})p_{z_c,h}$$
$$+\sum_{s_z\geq T_{z_{\min}}}R_{z+,a,h}(t,T_{z_{\min}},s_z)i_{z-}(s_z)p_{z_c,h}$$

$$P_{e-,a,h}(t+\delta t)=P_{e-,a,h}(t)+L_{z+,a,h}(t,T_{z_{\min}})i_{z-}(T_{z_{\min}})(1-p_{z_c,h})$$ Equation 1.11b
$$+\sum_{\substack{s_z=0,\\ s_z\neq T_{z_{\min}}}}^{T_{z_{\max}}}L_{z+,a,h}(t,s_z)i_{z-}(s_z)+R_{z+a,h}(t,T_R,T_{z_{\max}})$$

$$L_{z+,a,h}(t+\delta t,s_z+\delta s_z)=L_{z+,a,h}(t,s_z)(1-i_{z-}(s_z))(1-p_{z_c,h}) \qquad s_z=T_{z_{\min}}$$ Equation 1.12a

$$L_{z+,a,h}(t+\delta t,s_z+\delta s_z)=L_{z+,a,h}(t,s_z)(1-i_{z-}(s_z)) \qquad s_z\neq T_{z_{\min}}$$ Equation 1.12b

$$R_{z_p,a,h}(t+\delta t,s_r+\delta s_r)=R_{z_p,a,h}(t,s_r)+R_{z+,a,h}(t,s_r,T_{z_{\max}})$$ Equation 1.13a
$$+\sum_{s_z<T_{z_{\max}}}R_{z+,a,h}(t,s_r,s_z)i_{z-}(s_z)$$
$$s_r<T_{z_{\min}}$$

$$R_{z_p,a,h}(t+\delta t,s_r+\delta s_r)=R_{z_p,a,h}(t,s_r)+R_{z+,a,h}(t,s_r,T_{z_{\max}})(1-p_{z_c,h})$$
$$+\sum_{s_z\geq T_{z_{\min}}}R_{z+,a,h}(t,s_r,s_z)i_{z-}(s_z)(1-p_{z_c,h})$$ Equation 1.13b
$$+\sum_{s_z<T_{z_{\min}}}R_{z+,a,h}(t,s_r,s_z)i_{z-}(s_z)$$

$$s_r = T_{z_{\min}}$$

$$R_{z_p,a,h}(t+\delta t, s_r+\delta s_r) = R_{z_p,a,h}(t,s_r) + R_{z+,a,h}(t,s_r,T_{z_{\max}}) + R_{z+,a,h}(t,s_r,T_{z_{\min}}) i_{z-}(T_{z_{\min}})(1-p_{z_c,h}) + \sum_{s_z \neq T_{z_{\min}}} R_{z+,a,h}(t,s_r,s_z) i_{z-}(s_z) \quad \text{Equation 1.13c}$$

$$s_r > T_{z_{\min}}$$

$$R_{z+,a,h}(t,s_r+\delta s_r,s_z+\delta s_z) = R_{z+,a,h}(t,s_r,s_z)(1-i_{z-}(s_z))(1-p_{z_c,h}) \quad \text{Equation 1.14a}$$

$$s_r = T_{z_{\min}} \text{ and } T_{z_{\max}} > s_z \geq T_{z_{\min}}$$

$$\text{or } s_z = T_{z_{\min}} \text{ and } T_{z_{\max}} > s_r \geq T_{z_{\min}}$$

$$R_{z+,a,h}(t,s_r+\delta s_r,s_z+\delta s_z) = R_{z+,a,h}(t,s_r,s_z)(1-i_{z-}(s_z)) \quad \text{Equation 1.14b}$$

$$s_r < T_{z_{\min}} \text{ and } s_z < T_{z_{\max}}$$

$$\text{or } s_r = T_{z_{\min}} \text{ and } s_z < T_{z_{\min}}$$

$$\text{or } s_r > T_{z_{\min}} \text{ and } s_z \neq T_{z_{\min}} \text{ and } s_z < T_{z_{\max}}$$

$$V_{z+,a,h}(t+\delta t, s_z+\delta s_z) = V_{z+,a,h}(t,s_z)(1-i_{z-}(s_z)) \qquad s_z < T_{z_{\min}} \quad \text{Equation 1.15a}$$

$$V_{z+,a,h}(t+\delta t, s_z+\delta s_z) = V_{z+,a,h}(t,s_z)(1-i_{z-}(s_z)) + L_{z+,a,h}(t,s_z)(1-i_{z-}(s_z))p_{z_c} + \sum_{s_r \geq T_{z_{\min}}} R_{z+,a,h}(t,s_r,T_{z_{\min}})(1-i_{z-}(T_{z_{\min}}))p_{z_c,h} \quad \text{Equation 1.16a}$$

$$s_z = T_{z_{\min}}$$

$$V_{z+,a,h}(t+\delta t, s_z+\delta s_z) = V_{z+,a,h}(t,s_z)(1-i_{z-}(s_z)) + R_{z+,a,h}(t,T_{z_{\min}},s_z)(1-i_{z-}(s_z))p_{z_c,h} \quad \text{Equation 1.16b}$$

$$s_z > T_{z_{\min}}$$

Web Table 1: Definitions of the compartments and variables in the model.

Symbol	Definition
$L_{z-,a,h}(t)$	Number of miners of age *a* and HIV status *h* in the latent category at time *t*, not on IPT.
$L_{z+,a,h}(t,s_z)$	Number of miners of age *a* and HIV status *h* in the latent category at time *t*, who have been on IPT for duration s_z.
$P_{e+,a,h}(t)$	Number of miners of age *a* who have previously had IPT, cleared their infection and have not been reinfected since clearing their infection.
$P_{e-,a,h}(t)$	Number of miners of age *a* who have had IPT, have not cleared their infection and have not been reinfected during the previous two years
$R_{z-,a,h}(t,s_r)$	Number of miners of age *a* and HIV status *h* who have been reinfected for duration s_r at time *t*, who have never had IPT. The maximum time, T_R for which people can be in this reinfected category is 2 years.
$R_{z+,a,h}(t,s_r,s_z)$	Number of miners of age *a* and HIV status *h* who have been reinfected for duration s_r ($<T_R$) and have been on IPT for duration s_z at time *t*.
$R_{z_p,a,h}(t,s_r)$	Number of miners of age *a* and HIV status *h* who have been reinfected for duration s_r ($<T_R$) at time *t*, who have previously had IPT.
$E_{n,s,a,h}(t,s_o)$	Number of undetected cases of age *a,* HIV status *h* and smear status *s* who have had disease through (endogenous) reactivation for duration s_o at time *t,* if $s_o < T_{o_{\max}}$ (2 years). If $s_o = T_{o_{\max}}$, $E_{n,s,a,h}(t,s_o)$ represents the number of cases of age *a,* HIV status *h,* smear status *s* who have had disease through (endogenous) reactivation for at least time $T_{o_{\max}}$.
$E_{x,s,a,h}(t,s_o)$	Number of undetected cases of age *a,* HIV status *h* and smear status *s* who have had disease because of (exogenous) reinfection for duration s_o at time *t*, if $s_o < T_{o_{\max}}$. If $s_o = T_{o_{\max}}$, $E_{x,s,a,h}(t,s_o)$ represents the number of cases of age *a,* HIV status *h,* smear status *s* who have had disease because of (exogenous) reinfection for at least time $T_{o_{\max}}$ at time *t*
$F_{s,a,h}(t,s_f)$	Number of cases of smear status *s*, age *a,* HIV status *h* who have been detected ("found") for duration s_f at time *t* and have not yet started TB treatment.
$C_{a,h}(t,s_\tau)$	Number of cases of age *a,* HIV status *h* who have been on TB treatment for duration s_τ at time *t*.
$V_{z-,a,h}(t)$	Number of miners of age *a,* HIV status *h* who are in the recovered category at time *t* who are not on IPT.
$V_{z+,a,h}(t,s_z)$	Number of miners of age *a,* HIV status *h* who are in the recovered category at time *t* who have been taking IPT for duration s_z.
$M_{in,L,a,h}(t)$	Number of new employees at time *t,* who are of age *a,* HIV status *h* and in the latent category.
$M_{in,R,a,h}(t,s_r)$	Number of new employees joining the mining workforce at time *t* who are of age *a,* HIV status *h* and who have been reinfected for duration s_r
$M_{in,E_n,s,a,h}(t,s_o$	Number of new employees at time *t* who are of age *a,* HIV status *h* who have been experiencing disease because of endogenous reactivation for duration s_o, and currently have smear status *s*.
$M_{in,E_x,s,a,h}(t,s_o$	Number of new employees at time *t* of age *a,* HIV status *h* who have been experiencing disease following exogenous reinfection for duration s_o, and currently have smear status *s*.
$M_{in,V,a,h}(t)$	Number of new employees at time *t* of age *a,* HIV status *h* who have previously had TB, been treated and have not been reinfected since then.

Web Appendix 2: Transitions and parameters in the model

2.1 The force (or risk) of infection

The force (or risk) of infection in each cluster changes over time and is calculated using Equation 2.1 as the sum of the force of infection attributable to contact with individuals outside the mining population (λ_o) and that attributable to contact with miners within the same cluster ($\lambda_w(t)$):

$$\lambda(t) = \lambda_w(t) + \lambda_o \qquad \text{Equation 2.1}$$

λ_o was assumed to be 0.29%/year, calculated using the typical time spent with the outside community (7 hours or 100×7/168=4.2% of each week, found in the baseline prevalence survey) and the assumed force of infection in the community (7%/year), consistent with estimates among adolescents from a similar community(*1*).

The force of infection at time *t* attributable to contact with miners within the same cluster ($\lambda_w(t)$) is given in Equation 2.2 in terms of the effective contact rate (c_e) (defined as the average number of individuals effectively contacted by each infectious case), the total number of smear-negative and smear-positive individuals ($I_{s-}(t)$ and $I_{s+}(t)$ respectively), the size of the cluster ($N(t)$) and the relative infectiousness of smear-negative, compared to smear-positive cases (f). The latter equals 22%, consistent with molecular epidemiological data(*2*).

$$\lambda_w(t) = \frac{c_e(fI_{s-}(t) + I_{s+}(t))}{N(t)} \qquad \text{Equation 2.2}$$

Extending the definition used for acute infections(*3*), an effective contact is defined as one that is sufficient to lead to transmission if it occurs between an infectious individual and someone with either a “latent” infection or who has never been infected.

The total number of smear-positive individuals is given by the following equation:

$$I_{s+}(t) = \sum_{a}\sum_{h}\sum_{s_o=0}^{T_{o_{max}}}(E_{n,s+,a,h}(t,s_o) + E_{x,s+a,h}(t,s_o)) + \sum_{s_f=0}^{T_{f_{max}}} F_{s+,a,h}(t,s_f)$$

The equation for smear-negative cases is analogous.

Differences in the rates of case detection between clusters (section 2.4) means that, for a given value for the effective contact rate, the prevalence of infectious individuals and therefore, the annual risk of infection, differs between clusters.

In the base case model, the effective contact rate was chosen so that it led to an annual risk of infection (ARI) averaged across all intervention clusters before the start of the intervention of 20%/year. This value is consistent with studies among goldminers from the 1960s(*4*), when the TB prevalence was lower than that seen in recent years (i.e. 4.66 per 1000) which suggested that the ARI was at least 16-18%/year. Other studies, considering the relapse rates following TB treatment among HIV-positive and HIV-negatives have suggested that ARI was at least 10% per year(*5-6*). The effect of alternative values of 10% and 30%/year for the average ARI were explored in sensitivity analyses.

2.2 The rates of disease onset

2.2.1 Smear-negative TB

2.2.1.1 HIV-negative individuals

The overall age-specific rates of disease onset among HIV-negatives in the age groups <40 or ≥40 years in the absence of IPT are calculated as the weighted average of the rate for those with and without radiologically-confirmed silicosis (see below for how these are

estimated). Considering reactivation disease and in the absence of IPT, the expression is given by $d_{n,z-,a,h-}k_{si,a,h-}$ where the scaling factor $k_{si,a,h-}$ is given by the following equation:

$$k_{si,a,h-} = (1 - p_{si,a}) + p_{si,a}\rho_{si,h-}$$

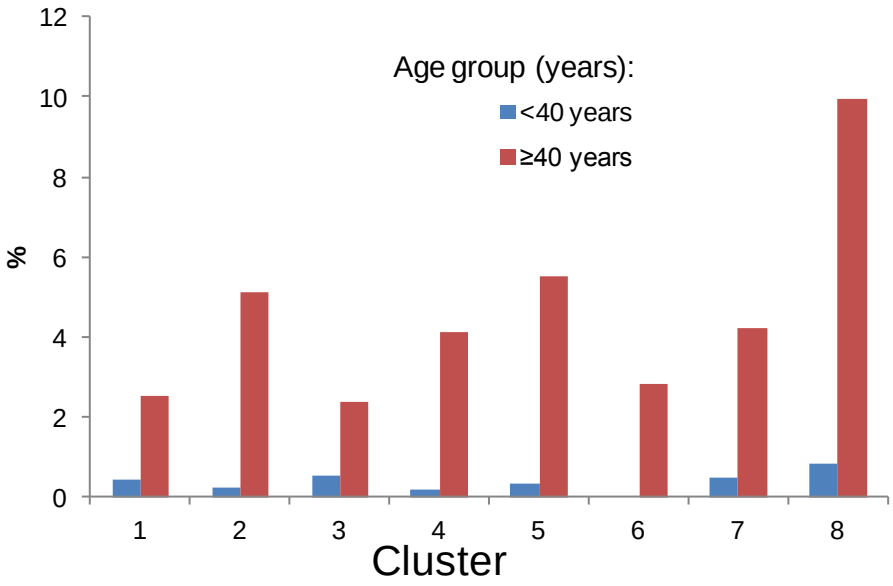
Here, $d_{n,a,h-}$ is the rate of onset of reactivation disease among miners in age group *a* without radiologically-confirmed silicosis, p_{si} is the prevalence of radiologically-confirmed silicosis, and $\rho_{si,h-}$ is the relative risk of developing TB disease among HIV-negative miners without radiologically-confirmed silicosis, compared to HIV-negative individuals without silicosis. The latter relative risk was assumed to be 2.6, which is consistent with findings from Corbett et al(*7*). The expression for the overall rate at which HIV-negative individuals experienced disease following reinfection is analogous. Since the prevalence of silicosis differs between clusters, the overall rates at which HIV-negatives experience disease through reactivation or following reinfection also differ by cluster.

The rate at which miners develop disease following reinfection in the absence of IPT (denoted by $d_{x,z-,a,h}(0)$ is assumed to be highest during the first year after reinfection, with that in the second year after reinfection differing by a constant factor (0.41) from that in the first year. This factor is based on findings from historical data suggesting a decrease in the rate of disease onset with time since first infection(*8-9*). There are no empirical data showing that reinfection-related disease follows this pattern, although the assumed relationship is plausible. For consistency with this relationship, the reactivation rate is assumed to be up to 13% of the rate of disease in the first year following reinfection. Since it is unlikely that individuals experience disease in the first few weeks after reinfection, we assume that disease can first occur two months after reinfection.

The rates of reactivation and disease onset following reinfection have been estimated for HIV-negative populations(*10-13*), but probably differ for mining populations, given exposure to silica dust, which may increase susceptibility to mycobacterial disease even among those

with sub-radiological silicosis(*7*). As the cumulative exposure to silica dust and susceptibility to disease probably increases with increasing age, the disease rates are assumed to depend on age (differing between those aged <40 and ≥40 years). The values for HIV-negative miners without radiologically-confirmed silicosis are estimated (along with other unknown parameters) by fitting model predictions of the incidence and prevalence outcomes from the intervention to the observed data.

IPT is assumed to provide 63% protection against disease ($\pi_{d,z+,ART-}$) (section 2.6.1). The rates of onset of reactivation disease among HIV-negative individuals who are on IPT was given by the expression $(1-\pi_{d,z+,ART-})d_{n,z-,a,h-}$.



Web Figure 2: Age-specific prevalence of radiologically-confirmed silicosis measured in the baseline prevalence survey

2.2.1.2 HIV-positive individuals

2.2.1.2.1 Rates of disease onset in the absence of IPT

The age-specific rates at which HIV-positive miners without radiologically-confirmed silicosis developed tuberculosis disease through reactivation in the absence of IPT at a given time *t* was calculated using the following equation as the weighted average of the rate of disease

onset among those who were not on ART and those on ART, accounting for the fact that ART was only provided for those with a CD4 count of <200 cells/mL:

$$d_{n,z-,a,h+}(t) = d_{n,z-,a,h-}k_{z-,ART,h}(t) \qquad \text{Equation 2.3}$$

where $k_{z-,ART,h}(t)$ is given by the following equation:

$$k_{z-,ART,h}(t) = p_{h,<200}g_{ART,<200}(t)\rho_{h,<200}(1-\pi_{z-,ART+}) + p_{h,<200}(1-g_{ART,<200}(t))\rho_{h,<200} + (1-p_{h<200})\rho_{h,\geq 200}$$

In these equations $p_{h,<200}$ is the proportion of HIV-positive miners with a CD4 count of <200 cells/mL (assumed to be 25% - see section 2.3.2.1), $g_{ART,<200}(t)$ is the ART coverage among those with a CD4 count of <200 cells/mL (changes over time and depends on the cluster – see section 2.3.2.2), $\pi_{z-,ART+}$ is the protection provided by ART against tuberculosis disease (65% - see section 2.3.3); $\rho_{h,<200}$ and $\rho_{h,\geq 200}$ are the relative risks of developing tuberculosis among those with a CD4 count of <200 cells/mL and ≥200 cells/mL respectively, which were assumed to be 17.0 and 5.9 respectively. These are consistent with estimates in (*14*).

Equation 2.3 was scaled up by a factor $k_{si,a,h+}$ to account for the age- and cluster-specific prevalence of silicosis and the relative risk of developing tuberculosis among silicotic HIV-positive individuals, compared to those without silicosis ($\rho_{si,h+}$), to give the average rate of onset of disease through reactivation among HIV-positives in the cluster. Based on Corbett et al(*7*) the relative risk of developing tuberculosis among HIV-positive individuals with radiological signs of silicosis, compared to those without silicosis was assumed to be 4.1 in the base case model.

The equations for the rates of disease following exogenous reinfection for HIV-positive individuals are analogous.

2.2.1.2.2 Rates of disease onset after the introduction of IPT

The age-specific rates at which HIV-positive miners without radiologically-confirmed silicosis developed tuberculosis disease through reactivation whilst they were on IPT at a given time *t* were calculated in an analogous way to those for miners who were not IPT (Equation 2.3), using the following equation as the weighted average of the rate of disease onset among those who were on both ART and IPT, and those who were just on IPT:

$$d_{n,z+,a,h+}(t) = d_{n,z-,a,h-}k_{z+,ART,h}(t) \qquad \text{Equation 2.4}$$

where $k_{z+,ART,h}(t)$ is given by the following equation:

$$k_{z+,ART,h}(t) = p_{h,<200}g_{ART,<200}(t)\rho_{h,<200}(1-\pi_{z+,ART+}) + p_{h,<200}(1-g_{ART,<200}(t))\rho_{h,<200}(1-\pi_{z+,ART-}) + (1-p_{h<200})\rho_{h,>200}(1-\pi_{z+,ART-})$$

The protection against disease resulting from being on both IPT and ART was assumed to be 82.5% (see section 2.6.1).

2.2.2 Rate of onset of smear-positive disease

It is generally thought that, if left untreated, most smear-negative tuberculosis cases will eventually become smear-positive, unless they die in the meantime. Relatively little is known about the rate at which this occurs, although it is probably faster for HIV-positive than for HIV-negative smear-negative cases, given their correspondingly shorter time to detection (section 2.4.2.2).

We estimated these rates for HIV-positive and HIV-negative miners (denoted using the symbols $o_{s+,h+}$ and $o_{s+,h-}$ respectively) using the following equations (see derivation in section 2.2.2.1):

$$o_{s+,h+} = \frac{p_{s+,h+}}{D_{s+,h+}(1-p_{s+,h+})} \qquad \text{Equation 2.5a}$$

$$o_{s+,h-} = \frac{p_{s+,h-}}{D_{s+,h-}(1-p_{s+,h-})}$$ Equation 2.5b

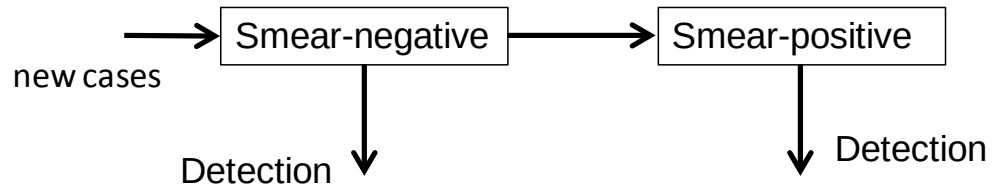
where $D_{s+,h+}$ and $D_{s+,h-}$ are the average times until smear-positive HIV+ and HIV- miners are detected and are based on data from Corbett et al(*15*) (7.5 and 51 weeks respectively – see section 2.4.2.1);

$p_{s+,h+}$ and $p_{s+,h-}$ are the proportions of HIV-positive and HIV-negative prevalent TB miners that were smear-positive, which are estimated to be 23.3% and 11.8% respectively, in the study of Corbett et al(*15*). These values are consistent with observed data from Thibela TB in the final prevalence survey (20% for all cases).

Substituting for $p_{s+,h+}$ = 0.118 and $D_{s+,h+}$ =7.5 weeks into Equation 2.5a and for $p_{s+,h-}$ = 0.233, $D_{s+,h-}$ = 51 weeks into Equation 2.5b leads to estimates of the rates at which HIV-positive and HIV-negative smear-negative cases become smear-positive of 1.76%/week and 0.6%/week respectively.

2.2.2.1 Deriving the expression for the rate of onset of smear-positive TB

Equation 2.5 was derived by considering data from Corbett et al(*15*) on TB cases, stratified by HIV status, who were smear-negative or smear-positive. The change in the composition of the TB cases in the study population can be described using the following diagram:



Web Figure 3: Diagram of a model describing the transition from smear-negative to smear-positive status

The rate of change in the proportion of HIV-positive TB cases that is smear-positive is given by the following equation:

$$\frac{dp_{s+,h+}}{dt} = o_{s+,h+}p_{s-,h+} - p_{s+,h+}r_{f,s+,h+,o} \qquad \text{Equation 2.6}$$

Here $p_{s-,h+}$ is the proportion of the HIV-positive TB cases that were smear-negative, $r_{f,s+,h+,o}$ is the overall rate at which HIV-positive smear-positive TB cases are detected and is approximated by 1/(estimated time to detection for smear-positives) or $1/D_{s+,h+.}$

The average (equilibrium) value for the proportion of HIV-positive TB cases that are smear-positive is obtained by setting Equation 2.6 to zero, which leads to the following relationship between the rate at which smear-negative cases become smear-positive, the rates at which smear-positive cases are detected and the proportion of TB cases that are smear-positive:

$$o_{s+,h+}p_{s-,h+} = p_{s+,h+}r_{f,s+,h+,o} \qquad \text{Equation 2.7}$$

After rearranging this equation, we obtain the following equation for the rate at which smear-negative individuals become smear-positive:

$$o_{s+,h+} = \frac{p_{s+,h+}r_{f,s+,h+,o}}{p_{s-,h+}} \qquad \text{Equation 2.8}$$

Using the result that $r_{f,s+,h+,o}$ approximately equals 1/(estimated time to detection for smear-positives) or $1/D_{s+,h+,}$ and that the proportion of HIV-positive TB cases that are smear-negative equals 1- proportion of HIV-positive TB cases that are smear-positive (i.e. ($p_{s-,h+}$ =1- $p_{s+,h+}$) leads to Equation 2.5a:

$$o_{s+,h+} = \frac{p_{s+,h+}}{D_{s+,h+}(1-p_{s+,h+})}$$

2.3 HIV and ART

2.3.1 Prevalence of HIV

The prevalence of HIV in the workforce is assumed to be 30% for each age group, as found by a study carried out during 2000-2001(*16*). There are no recent data on the prevalence of HIV, although it is unlikely to have changed substantially since then. We therefore assumed

that it remained unchanged since 2000 and explore the effect of assuming alternative plausible values for the prevalence, ranging between 20% and 40%.

2.3.2 ART coverage

2.3.2.1 The proportion of the workforce in a cluster that was eligible for ART

For simplicity, only individuals with a CD4 count of <200 cells/mL were assumed to be eligible for ART from the introduction of ART until the end of the Thibela study. In reality, the threshold at which individuals started ART varied between clusters (Grant, *personal comm*), e.g. a CD4 threshold of <200 cells/ml or WHO stage 4 for company B, and three criteria (WHO stage 4 and any CD4 count, WHO stage 3 (pulmonary TB) and CD4<350cells/mL; and WHO stage 1 or 2 and CD4<250 cells/mL) for company A. However, the proportion of individuals in the workforce in the affected clusters who would have been in WHO stage 3 or 4 was relatively small and is unlikely to affect coverage estimates. Including changes to the threshold for initiating ART is also unlikely to affect analyses considering time frame of the Thibela intervention since the threshold increased to 350 cells/mL in 2010 in South Africa, by which time, the final prevalence survey in most of the clusters would have been completed.

This proportion of HIV-positive miners with a CD4 count of <200 cells/mL is assumed to be 25%, consistent with data from Williams et al(*14, 17*). For simplicity, this proportion is assumed to have remained unchanged over time, although in reality, it may have changed slightly as the immune-suppression among who adhered to ART improved over time. Such changes are probably negligible during the time course of the Thibela study, given the low proportions of the clusters on ART during the early 2000s.

2.3.2.2 Calculating the ART coverage among those eligible

The coverage of ART among those with an initial CD4 count of <200 cells/mL was calculated assuming that ART was introduced in January 2003 in company A and in January 2004 in companies B and C. This is consistent with observations that ART was first introduced by Company A in November 2002; the year of introduction in companies B and C are unknown, but is thought to be later than 2002.

The ART coverage at a given time t after the introduction of ART at time t_0 was calculated using data from the baseline and final prevalence surveys on the proportion of miners in each cluster who reported to have ever taken ART (Figure 3A in the main text) using the following equation:

$$c_{ART}(t) = d(t - t_0)^b \qquad t < T_{FPS} \qquad \text{Equation 2.9}$$

$$c_{ART}(t) = d(T_{FPS} - t_0)^b \qquad t \geq T_{FPS}$$

Here. d and b are constant terms, depending on the cluster, calculated so that the coverage during the baseline and final prevalence surveys among those eligible equalled that estimated, based on the assumed prevalence of HIV and the proportion of HIV-positive individuals who had a CD4 count of <200 cells/mL; T_{FPS} is the time of the final prevalence survey.

Denoting the time of the baseline prevalence survey by T_{BPS}, we obtain the following equations for d and b:

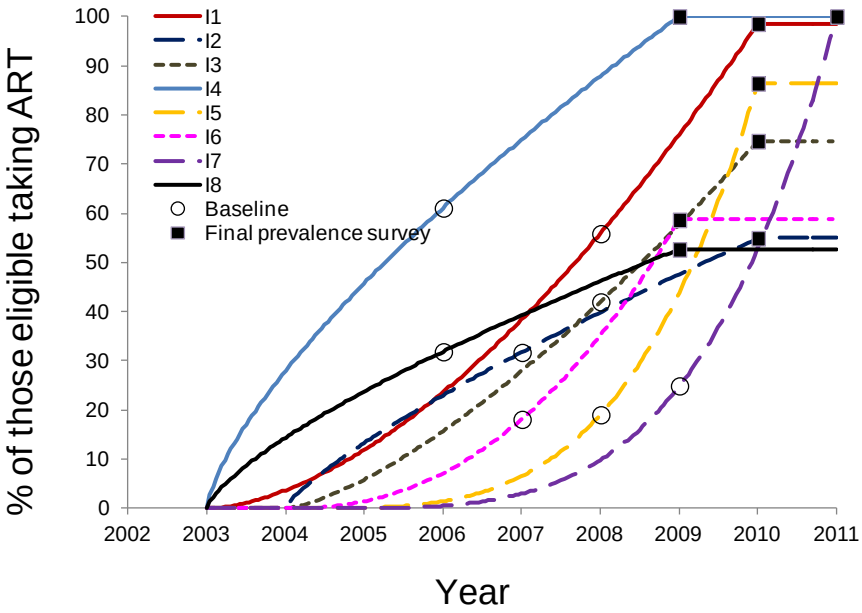
$$b = \frac{\ln(c_{ART}(T_{BPS}))}{\ln(T_{BPS} - t_0) - \ln(T_{FPS} - t_0)} \qquad \text{Equation 2.10}$$

$$d = \frac{c_{ART}(T_{BPS})}{(T_{FPS} - t_0)^b} \qquad \text{Equation 2.11}$$

The ART coverage among those eligible in these equations in the baseline and final prevalence surveys ($c_{ART}(T_{BPS})$ and $c_{ART}(T_{FPS})$ respectively) were calculated as the overall proportion of miners who had previously taken ART in the cluster at these times ($\tilde{c}_{ART}(T_{BPS})$ and $\tilde{c}_{ART}(T_{FPS})$), divided by the HIV prevalence (h_+) and the proportion of HIV-positive individuals with a CD4 count of less than 200 cells/mL:

$$c_{ART}(T_{BPS}) = \frac{\tilde{c}_{ART}(T_{BPS})}{h_+ p_{h,<200}} \qquad \text{Equation 2.12}$$

$$c_{ART}(T_{FPS}) = \frac{\tilde{c}_{ART}(T_{FPS})}{h_+ p_{h,<200}} \qquad \text{Equation 2.13}$$



Web Figure 4: Predicted coverage of ART among those with an initial CD4 count of <200 cells/mL, obtained assuming that the ARVs were introduced in the year January 2003 in company A and in January 2004 in companies B and C. The circles and boxes represent estimates for the baseline and final prevalence surveys respectively. The solid, dashed and dotted lines reflect clusters in companies A, B and C respectively. The coverage in a given year, *y*, is calculated using the expression $a(y-y_0)^b$, where y_0 is the year in which ART is introduced and *a* and *b* are constant terms (depending on the cluster). The constant terms are calculated so that the coverage at the time of baseline and final prevalence surveys among those eligible is similar to that estimated. The coverage is assumed to have remained unchanged after the final prevalence survey.

2.3.3 Protection provided by ART against tuberculosis disease

The protection against tuberculosis disease (through reactivation or following reinfection) for miners who are on ARVs but are not currently taking IPT ($\pi_{d,z-,ART+}$) is assumed to be 65%, as suggested by a recent review(*18*) considering 11 studies. For simplicity, the protection is

assumed to be independent of CD4 count, given the width of the confidence interval for the hazard ratio for those with a CD4 count of >200 cells/mL or <200 cells/mL for these studies, shown below:

Web Table 2: Findings of the hazard ratio (HR) for developing tuberculosis, according to CD4 count in a recent review of 11 studies (*18*)

CD4 Counts (cells/mL)	HR (95% CI)	Number of studies
<200	0.16 (0.07-0.36)	2
200-350	0.34 (0.19-0.60)	4
>350	0.43 (0.30-0.63)	3
All	0.35 (0.28-0.44)	11

2.4 Case detection rates

2.4.1 Routine medical exam

It is assumed that at the routine medical exam, the health services of each company take a chest X-ray from all presenting miners, follow up a proportion (65% - see below) of those with an abnormal X-ray and then take a sputum smear or culture for those followed up. Assuming that it takes roughly 15 months (64 weeks) to call up 95% of the mining workforce in a given cluster for the routine medical examination (*19*), the weekly rate at which cases with smear status *s* are detected through this screen was calculated using the equation:

$$r_{f,s,h,y} = 1-(1-0.95p_{f,s,h,y})^{1/64} \qquad \text{Equation 2.14}$$

where $p_{f,s,h,y}$ is the proportion of cases of smear status *s*, HIV status *h*, attending the routine medical examination that are detected. This proportion is 0.25 for smear-positive cases for all companies and 0.137 and 0.014 for smear-negative cases for companies A and B/C respectively (see below). These values lead to values of 0.423%/week for all companies for the rate at which smear-positive cases are detected through the routine medical exam, and to values of 0.217% and 0.020%/week for smear-negatives for companies A and B/C respectively.

The above calculations assume that the rates at which these checks occur are independent of HIV status, which is plausible.

2.4.1.1 The proportion of cases attending the routine medical exam that are detected

The proportion of cases with smear status *s* attending the routine medical exam that are detected, $p_{f,s,h,y,}$ is calculated using the following equation:

$$p_{f,s,h,y} =$$
(sensitivity of X-ray for detecting TB of smear status s)
×
(proportion of cases that are investigated further following the X-ray)
×
(sensitivity of the method used in further investigations) Equation 2.15

X-ray is assumed to have a sensitivity of 21% and 38.5% for detecting smear-negative and smear-positive TB respectively, based on findings from Lewis et al(*16*), which considered miners who would have been similar to those in the study population. For simplicity, the sensitivity is assumed to be independent of HIV status.

For company A, 65% of TB cases are assumed to be investigated further based on X-ray findings. This is derived from findings in Churchyard et al(*19*) that only 43% of those with an abnormal radiology at the routine screen were followed up. This low estimate has been attributed to poor referral practices and miners not returning for subsequent investigations(*19*). This percentage probably underestimates the proportion of TB cases that are investigated given that TB cases with an abnormal X-ray are more likely to be symptomatic and their X-rays are more likely to be suggestive of TB than are others with an abnormal X-ray. These proportions could well also depend on smear-status. Since the exact value is unknown, for simplicity, it is taken to be 50% higher than that seen, i.e. 1.5×0.43×100=65%.

In company A, it is assumed that the isolates from most cases who return for subsequent investigation are cultured and therefore the sensitivity used in Equation 2.15 is 100%.

There are no data on the proportion of TB cases in companies B and C that are investigated further for TB, based on X-ray findings. Since it is unlikely to differ from that in company A, it is assumed to be 65%.

Companies B and C are assumed to carry out smear microscopy on all of those who are investigated further and should therefore identify most of the smear-positive cases that are followed up. They are assumed to culture isolates only from those who have had previous TB, which is assumed to be 10% of the TB cases (based on the average proportion of TB cases in the baseline prevalence survey who were on TB treatment) which means that only 10% of smear-negative cases presenting at the routine medical exam would be detected.

2.4.2 Rates at which cases are detected through passive presentation

2.4.2.1 Smear-positive cases - both companies

For simplicity, we assume that smear-positive cases are detected through presenting passively at the same rates in all companies, since each company uses the same method for detecting smear-positive cases (i.e. smear microscopy). This rate for TB cases of HIV status *h* was estimated as the difference between the overall rate at which smear-positive cases of HIV status *h* ($1/D_{s+,h}$) are detected and the rate at which smear-positive cases are detected in the routine medical exam ($r_{f,s+,h,y}$), as follows:

$$r_{f,s+,h,p} = 1/D_{s+,h} - r_{d,s+,h+,a} \qquad \text{Equation 2.16}$$

Here $D_{s+,h}$ is the average time to detection of smear-positive disease for miners of HIV status *h*, and was inferred from data in Corbett et al(*15*) as follows. That study estimated the average duration or time to detection of smear-positive TB by HIV status in clusters in company A during July 2000-January 2001 as the ratio between the observed incidence of smear-positive cases and the prevalence of smear-positive TB. This ratio was reported to be 8.84 weeks and 59.8 weeks for HIV-positive and HIV-negative individuals respectively.

The calculation accounted for outmigration and death, but probably overestimated the true time to detection, since the observed incidence of smear-positive TB was calculated as the number of cases who started TB treatment following detection and therefore did not account for the initial loss to follow-up among those detected. Assuming that, as observed in Thibela TB, the initial loss to follow-up was 15% among those who did not die or leave the workforce following detection leads to estimates of the average time to detection of smear-positive TB of 7.5 and 51 weeks for HIV positives and HIV-negatives respectively ($D_{s+,h+}$ and $D_{s+,h-}$). Substituting these values for $D_{s+,h+}$ and $D_{s+,h-}$ into Equation 2.16 leads to values of 0.1295/week and 0.0155/week for the rates at which HIV-positive and HIV-negative smear-positive cases are detected through passive presentation.

2.4.2.2 Smear-negative cases – company A

The rate at which HIV-positive smear-negative cases in company A are detected through passive presentation was estimated using an analogous expression to that used for HIV-positive smear-positive cases (Equation 2.16), and also accounting for the fact that smear-negative cases become smear-positive (at a rate $o_{s+,h+}$):

$$r_{f,s-,h,p} = 1/D_{s-,h} - r_{f,s-,h+,y} - o_{s+,h+} \qquad \text{Equation 2.17}$$

Here $D_{s-,h}$ is the average time to detection of smear-negative disease for individuals of HIV status *h.*

$D_{s-,h+}$ and $D_{s-,h-}$ were inferred from data in Corbett et al(*15*), using the same approach used for calculating the corresponding value for smear-positive cases (see above). As Corbett et al(*15*) did not explicitly present the observed incidence or prevalence of smear-negative TB, we inferred these statistics from the available published data (*15*). The observed incidence by HIV status was calculated as the difference between the observed incidence of all diagnosed TB and smear-positive TB, stratified by HIV status, where all diagnosed TB cases

include both cases with microbiological confirmation and those who were started on TB treatment without microbiological confirmation (3.4 and 0.67 per 100 years of follow-up for HIV-negative and HIV-positive individuals respectively – see Web Table 3). The prevalence of smear-negative TB for HIV-positive and HIV-negative individuals was calculated similarly (3.36% and 1.75% respectively - Web Table 3).

Taking the ratio between the prevalence and observed incidence of smear-negative TB leads to estimates of an observed time to detection of smear-negative TB of 51 and 136 weeks for HIV-positive and HIV-negative individuals respectively. Assuming that, as observed during Thibela TB, the initial loss to follow-up was 40% among the detected smear-negative TB cases in company A who did not leave the mines or die after detection leads to estimates of the time to detection of smear-negative cases of 31 weeks and 82 weeks (i.e. $D_{s-,h+}$ and $D_{s-,h-,}$) for HIV-positive and HIV-negative individuals respectively.

These values for $D_{s-,h+}$ and $D_{s-,h-,}$ together with the corresponding estimated rates at which smear-negative cases become smear-positive ($o_{s+,h+}$ and $o_{s+,h-}$) described in section 2.2.2 into Equation 2.17 leads to the following estimated rates at which HIV-positive and HIV-negative smear-negative cases are detected through passive presentation:

$$r_{f,s-,h+,p} = 1/31 - 0.00217 - 0.0176 = 0.0124 \text{ per week} \qquad \text{Equation 2.18}$$

$$r_{f,s-,h-,p} = 1/82 - 0.00217 - 0.006 = 0.0041 \text{ per week} \qquad \text{Equation 2.19}$$

2.4.2.3 Smear-negative cases – companies B/C

There are no data on the rate at which smear-negative cases are detected in companies B and C, although it is known that cases are detected mainly using smear-microscopy and, as for company A, a proportion are referred for TB treatment without microbiological confirmation.

Since the estimated rate at which smear-negative cases are detected through the routine medical exam in companies B and C is less than that for company A, the rates at which smear-negative cases are detected through passive presentation in companies B and C could well be higher than for company A, i.e. cases are increasingly likely to present passively if they are missed in the routine medical exam.

For simplicity, the rates at which smear-negative cases are detected through passive presentation, by HIV status in companies B and C, are calculated so that the overall rate at which they are detected equals the overall rate at which smear-negatives are detected in company A. The overall rate is calculated as the sum of the case detection rate resulting from the routine medical exam and passive presentation. The rates at which smear-negatives were detected through passive presentation in companies B and C was then calculated as 1.44%/week and 0.61%/week for HIV-positive and HIV-negative smear-negative TB cases respectively.

2.4.3 Initial screen on joining the workforce

The procedure for screening miners when they join the workforce is assumed to be identical to that used for screening miners at the routine medical exam. The proportion of cases that are detected when joining the mines then equals the proportion detected at the routine medical exam and differs between companies and according to smear status.

It is recognised that, in reality, contractors are not screened for TB either when joining the workforce or routinely and therefore the model may be overestimating the proportion of miners that are identified through these screening processes. However, this limitation is small, given the small proportion of mining employees that are contractors. In sensitivity analyses, we explore the effect of different assumptions about the prevalence of culture-positive TB among miners joining the workforce (section 2.7.4)

Web Table 3: The average "observed" and estimated time to detection of smear-positive and smear-negative TB, as calculated by dividing the corresponding average prevalence by either the observed or estimated incidence of smear-negative TB using data from Corbett et al(*15*). The incidence is calculated considering cases who either self- presented to the health services or were detected through the routine medical exam. The values for smear-positives (row I) which were presented in Corbett et al are reproduced here for convenience. All diagnosed cases are defined as all those who started TB treatment, irrespective of microbiological confirmation.

	Definition	**HIV+**	**HIV-**
Prevalence (%)	A. Culture-positive (all diagnosed)	3.8	2.3
	B. Smear-positive	0.44	0.55
	C. Smear negative (estimated as A-B)	3.36	1.75
Observed incidence (per 100 PYFU)	D. Culture-positive (all diagnosed)	6	1.15
	E. Smear-positive	2.6	0.48
	F. Smear negative (estimated as D-E)	3.4	0.67
Estimated number of detected cases per 100 PYFU	G. Smear-positive (=E/(1-initial loss to follow-up (0.15)))	3.06	0.56
	H. Smear negative (=F/(1-initial loss to follow-up (0.4)))	5.67	1.12
I. Observed time to detection of sm+ cases, (B/E)		0.44/2.6=0.17 yrs or 8.8 weeks	0.55/0.48=1.15 yrs or 59.6 weeks
J. Observed time to detection of sm- cases (C/F)		3.36/3.4 =0.99 yrs or 51.4 weeks	1.75/0.67 = 2.6 yrs or 135.8 weeks
K. Estimated time to detection of sm+ cases, after correcting the observed time to detection for initial loss to follow-up, $D_{s+,h+}$ and $D_{s+,h-}$ (B/G)		0.44/3.06 =0.14 yrs=7.48 weeks	0.55/0.56 =0.97 yrs=50.65 weeks
L. Estimated time to detection of sm- cases, after correcting the observed time to detection for initial loss to follow-up, $D_{s-,h+}$ and $D_{s-,h-}$ (C/H)		3.36/5.67 =0.59 yrs=30.8 weeks	1.75/1.12 = 1.57 yrs=81.5 weeks

Abbreviations: sm+: smear-positive; sm-: smear-negative, PYFU: per year of follow-up.

.

2.5 Initial loss to follow-up and time to TB treatment

The initial loss to follow-up among those referred for TB treatment depends on both smear status and company as observed in Thibela TB (Web Table 4). Note that for the purposes of the model, initial loss to follow-up is defined as the proportion of individuals who have not started treatment 6 months after detection, as data from the trial suggest that few cases start treatment more than 6 months after detection.

Web Table 4: Summary of the values for initial loss to follow-up used in the model

Company	**Smear-negative**	**Smear-positive**
A	40%	15%
B	55%	30%
C	55%	30%
Average	50%	28%

The rates at which miners start treatment in the model depend on smear status, company and time since treatment and were calculated such that the proportion of individuals who had not started treatment at month *i* (*i*=1,...,6) after detection matched that observed in Thibela TB (see Web Table 5 and Web Table 6). Miners who had not started treatment within 6 months were returned to the undetected categories.

These rates were calculated as the values for $\tau(s_f)$ satisfying the following equations:

$$u(s_f + 1) = u(s_f) - \tau(s_f)u(s_f)$$

$$u(4i) = p_{tr\_}(i)$$

where: $\tau(s_f)$ is the rate at which cases start TB treatment in week s_f after detection (assumed to be constant in each month); $u(s_f)$ is the estimated proportion of those detected who are still untreated s_f weeks after detection; $p_{tr\_}(i)$ is the observed proportion of those detected who are still untreated at month *i* after detection (see Web Table 5 and Web Table 6).

Web Table 5: The observed proportion of cases who were not on treatment by the end of each month after detection

	Company					
	A		**B**		**C**	
Month (wk)	**Sm-**	**Sm+**	**Sm-**	**Sm+**	**Sm-**	**Sm+**
1 (4)	0.75	0.27	0.86	0.46	0.79	0.39
2 (8)	0.54	0.18	0.71	0.36	0.71	0.34
3 (12)	0.47	0.16	0.64	0.34	0.62	0.33
4 (16)	0.45	0.16	0.59	0.32	0.58	0.31
5 (20)	0.43	0.15	0.55	0.31	0.57	0.30
6 (24)	0.40	0.15	0.55	0.30	0.55	0.30

Web Table 6: The assumed weekly rates at which individuals started treatment in each month following detection

	Company					
	A		B		C	
Month	**Sm-**	**Sm+**	**Sm-**	**Sm+**	**Sm-**	**Sm+**
1	0.070	0.279	0.038	0.176	0.058	0.212
2	0.079	0.098	0.046	0.058	0.025	0.026
3	0.032	0.022	0.028	0.018	0.034	0.011
4	0.013	0.009	0.020	0.016	0.017	0.019
5	0.012	0.006	0.018	0.005	0.004	0.021
6	0.022	0.000	0.004	0.005	0.015	0.003

As shown in Web Table 7, the average time to treatment among those who start TB treatment is 7-8 weeks for smear-negatives and 3-4 weeks for smear-positive cases.

Web Table 7: Summary of the average times to treatment (weeks) among those who eventually start treatment used in the model

Company	Smear-negative	Smear-positive
A	7.4	3.1
B	8.6	4.1
C	8.1	4.0

2.6 IPT protection and coverage and case detection

2.6.1 Protection provided by IPT

IPT was assumed to provide 63% protection against disease through reactivation or following reinfection for both HIV-negative and HIV-positive individuals who are not on ARVs. This level of protection is based on findings from the individual-level analyses in Thibela TB(*20*); as it is based on all participants in Thibela TB, irrespective of tuberculin status(*20*), it accounts for the minimal benefit that tuberculin-negative individuals derive from IPT(*21*). It is also consistent with estimates from Comstock (1962)(*22*) which found a 68% reduction in TB incidence among those in the intervention arm compared to the control arm.

For simplicity, the assumed protection is independent of silicosis status. For those currently taking IPT, it also independent of the duration for which they have been taking IPT, which is consistent with findings from the individual-level estimates(*20*).

Those who are both on ART and IPT are assumed to have 82.5% protection against disease. This protection is calculated assuming that IPT provides an additional 50% protection to that provided by ART and that it is independent of CD4 count, which is consistent with the following:

1. Data from Botswana, which showed a 50% additional benefit of taking ART to taking IPT alone(*21*). However, this may underestimate the additional effect since it was calculated by comparing those eligible for ART against those ineligible, and those eligible for ART had a low CD4 count and were therefore at increased risk of TB(*23*).
2. Data from a recent trial in South Africa(*24*) which reported a 37% reduction in TB incidence (calculated over a 4 year period) for those taking both IPT and ART compared to ART alone. The additional protection provided by IPT while taking ART was about 50% (Hazard ratio of 0.52, 95%CI: 0.27-1.01)
3. Golub et al(*25*), which found an adjusted hazard ratio of 0.11 (95% CI: 0.02-0.78) for those receiving both ART and IPT in a cohort of HIV-positive patients in clinics in South Africa.
4. Golub et al(*26*) which found a 76% reduction (p<0.001) in the TB risk among those receiving both ART and IPT in a cohort of HIV-positive patients in clinics in Rio.

2.6.2 IPT coverage

Estimates of the rates at which individuals started and stopped IPT were based on observed data (see below). As shown in Web Figure 6, the assumed proportion of each cluster that was on IPT was generally consistent with that observed.

2.6.2.1 The rate at which non-diseased miners start IPT in each cluster

The rate at which non-diseased individuals of age *a* started IPT in each month t ($i_{z+,a}(t)$) was calculated by dividing the observed number of individuals in that age group that had started IPT in that month ($O_{z+,a}(t)$) by the estimated number of individuals in the same age group

who would have been able to start IPT in month *t*, had they volunteered to do so ($N_{e,a}(t)$), i.e. using the following equation:

$$i_{z+,a}(t) = \frac{O_{z+,a}(t)}{N_{e,a}(t)}$$

Equation 2.20

It was assumed that all individuals in the cluster, apart from those on TB treatment, were equally likely to volunteer to start IPT. The number of individuals of age *a* who would have been able to start IPT at the start of the intervention ($N_{e,a}(0)$), had they volunteered to do so, was then approximated by the following expression:

(The number of individuals of age *a* in the population at the start of the intervention

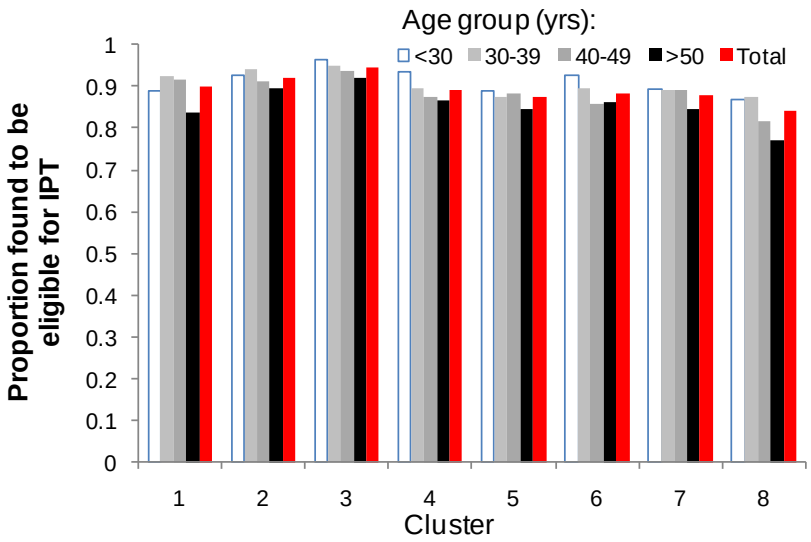
×

(The proportion of individuals of age *a* who were not on TB at the start of the intervention, as implied by the baseline prevalence survey)

×

(Proportion of the individuals of age *a* who consented for IPT during the entire enrolment period who were found to be eligible for IPT)

As shown in Web Figure 5, the proportion of the individuals who consented who were found to be eligible for IPT was age-dependent and differed between clusters.



Web Figure 5: Observed proportion of individuals in different age groups who consented for IPT who were subsequently found to be eligible for IPT.

The number of individuals of age *a* who would have been able to start IPT in each month after the start of the intervention ($N_{e,a}(t)$) was calculated assuming that it equalled the number observed in the previous month $N_{e,a}(t\text{-}1)$ after adding the observed number of individuals joining the cluster in that month ($M_{in,a}(t\text{-}1)$) and subtracting the observed numbers who had left the cluster $M_{out,a}(t\text{-}1)$ and the number who had started IPT that month ($O_{z+,a}(t\text{-}1)$), using the following equation:

$$N_{e,a}(t) = N_{e,a}(t-1) - M_{out,a}(t-1) + M_{in,a}(t-1) - O_{z+,a}(t-1) \qquad \text{Equation 2.21}$$

The above calculations of the rate at which those in the non-diseased (i.e. reinfected, latent and "recovered") compartments who started IPT each month implicitly assumes that the number of individuals who would have been found to be ineligible for IPT remained unchanged over time.

2.6.2.2 IPT retention

The assumed rate at which miners who were on IPT stopped IPT differed between clusters and, for simplicity, it was the same for all age groups. In general, the greatest proportion of those who stopped taking IPT did so within one month of starting IPT(*20*).

We therefore assumed that the rate at which miners stopped taking IPT was constant during the first month after they started IPT, and constant during the subsequent months. These rates were calculated using the age-specific monthly rates at which miners stopped taking IPT, averaged over the corresponding months and over all age groups. The age-specific rate for a given month *t* ($i_{z\text{-},a}(t)$) was calculated from the observed proportion of individuals in age group *a* that was still on IPT at the end of that month ($p_{z+,t,,a}$) and the average out-migration rate ($\overline{m}_a$), as follows:

$$i_{z-,a}(t) = 1 - \overline{m}_a - \frac{p_{z+,t,a}}{p_{z+,t-1,a}}$$

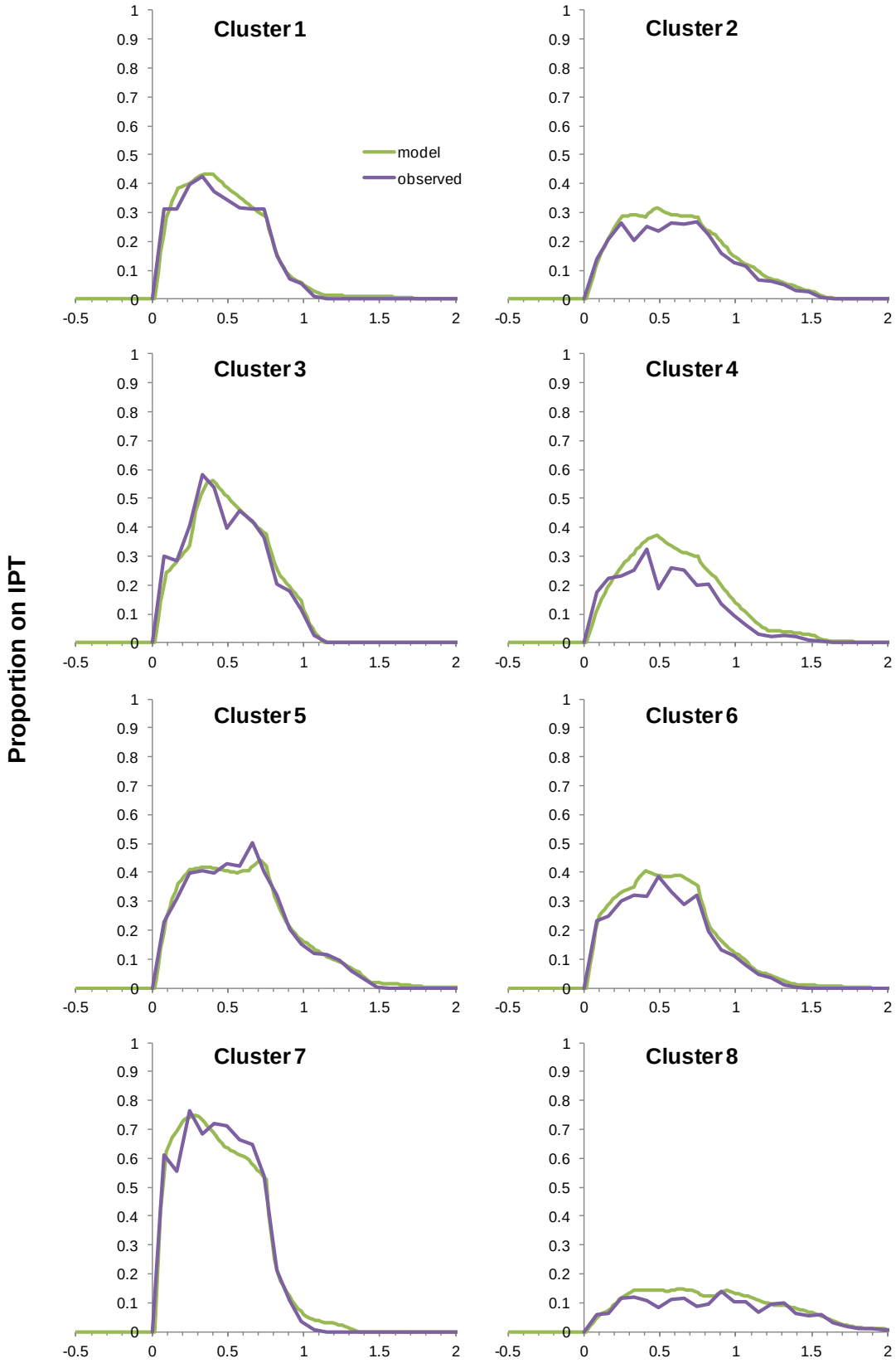
Equation 2.22

For a few clusters, the age-specific recruitment and retention rates calculated in the way described above led to slight under or overestimates (absolute difference of ~5%) of the maximum proportion of the cluster that was on IPT, due to slight differences between the actual migration rates in some months and the assumed average value. For those clusters, the rates at which miners stopped IPT after the first month was increased or decreased by small increments (1%) until the assumed proportion of individuals that were on IPT were consistent with observed values.

2.6.3 Case detection on recruitment into the intervention

Miners volunteering for IPT were screened for TB disease using a symptom questionnaire and chest X-ray(*27*). A single sputum specimen was taken for microscopy and culture from miners with symptoms or a chest X-ray which were suggestive of the miners having active TB and the miners were then sent for further investigation by the mine health service. The sensitivity of symptom screening carried out along with chest X-ray was about 49%(*16*). It is assumed that only cases which have not yet been detected came forward for the intervention, i.e. those who are either undergoing TB treatment or have already been detected but are awaiting TB treatment would not volunteer for IPT.

The monthly rates at which cases were detected in each cluster when consenting for IPT were calculated by dividing the observed number of culture-positive cases that were identified by the intervention when volunteering for IPT by the estimated average numbers of prevalent undetected culture-positive TB cases. The latter was approximated by multiplying the size of the cluster at the start of the intervention by the prevalence of culture-positive TB cases measured in the final prevalence survey and the average proportion of prevalent cases in the model that were undetected.



Web Figure 6: Comparison between the observed and assumed proportion of each cluster that was on IPT in each cluster

2.7 Demographic assumptions

2.7.1 Population size

We assumed that the age-specific number of miners in each cluster remained constant over time, which is generally consistent with observed data. However, miners were assumed to die or migrate out of each cluster at rates dependent on disease status. To ensure that the population remained constant over time, the number of miners in each of the four age groups (<30, 30-39, 40-49 and ≥50 years) joining each cluster each week was calculated to equal the total number of miners who either died or outmigrated each week (see below).

2.7.2 Mortality

Mortality for miners without TB disease was not explicitly modelled, as the outmigration rates calculated from the Human Resources database (see below) were interpretable as the combined mortality and out-migration rates. However, the mortality rates for TB cases were modelled and were assumed to be the same for all age groups, but differed between miners who had undetected TB disease and those who had started TB treatment.

The mortality rates for HIV-negative TB cases before they started TB treatment was assumed to be 1%/month for smear-positive and 0.2%/month for smear-negative tuberculosis cases. These were consistent with the average 10 year case-fatality rate reported in a systematic review of the case-fatality of untreated TB among HIV-negatives(*28*) which found a 10 year case-fatality rate of 53-86% (weighted average of 70%) and 20% for smear-positive and smear-negative cases respectively.

The mortality rates for HIV-positive TB cases before starting TB treatment is assumed to be 5%/month irrespective of smear status. There are no direct estimates of these mortality rates, as they are either typically calculated using data from cases who have already started TB treatment and are therefore probably underestimates, or accurate data on the

denominator (HIV/CD4 strata/person years of follow-up) are unavailable. In the absence of improved data, we use arbitrary (but plausible) values of approximately four times the mortality rate reported during the first 6 months after the start of treatment in one of the mining companies(*29*).

The mortality rates once tuberculosis cases start tuberculosis treatment are assumed to be 1.3%/month for HIV-positives and 0.13%/month for HIV-negatives and identical for smear-positive and smear-negative cases, which is consistent with the mortality rate reported during the first 6 months after the start of treatment in one of the mining companies (10% and 1% for HIV-positives and HIV-negatives respectively)(*29*). The absence of a difference by smear status, in contrast with those in studies conducted outside the mining community (*30*) may be due to mining health systems responding more actively to smear-positive cases than to smear-negative cases.

2.7.3 Out-migration

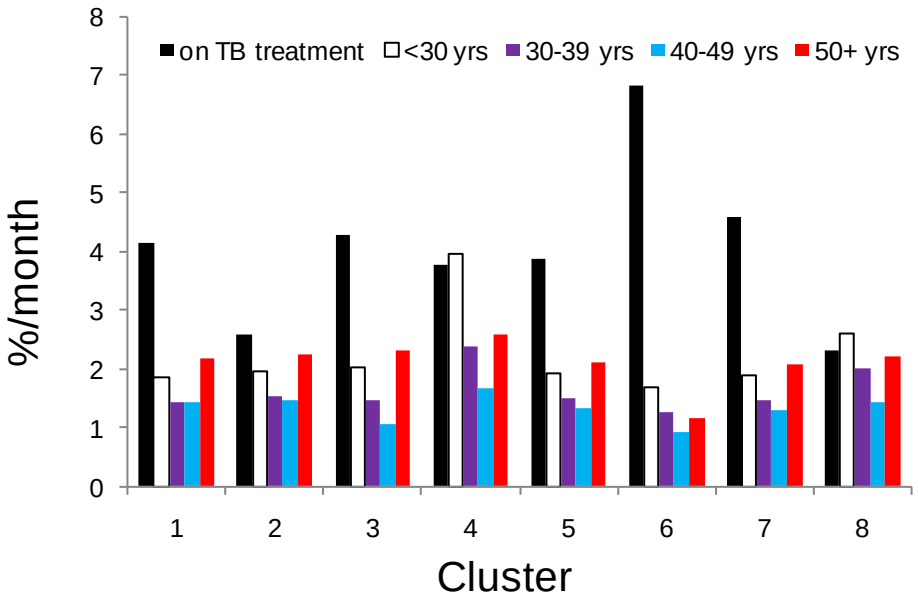
For simplicity, the model does not include movement between treatment and control arms, given the low levels at which this occurred. For example, considering the time period between the main enrolment and primary measurement periods, 0.46% (132/28,827) of miners moved from the control to the intervention arm and 0.34% (129/37,577) of miners moved from the intervention to the control arm.

Web Figure 7 summarizes the rates at which individuals in each cluster left the model as a result of leaving the workforce or non-specific mortality. These were assumed to be constant over time and differed between TB cases who had started treatment and all other individuals. The out-migration rates for the former were based on observed data and were assumed to be independent of age; those for all miners who were not on TB treatment were assumed to be age-dependent.

The out-migration rates for miners who were not on TB treatment were derived using data from the Human Resources department for each company, which provided the monthly numbers of miners in each age group who had joined, left or had been present in each cluster since the previous month. The reasons for why miners had left a given cluster are not provided and could therefore include death. These data were first used to calculate the age-specific rate at which miners left the workforce in each cluster each month due to death or other reasons during the study period, using the ratio between the number of miners that left the workforce each month in each of the four age groups (<30, 30-39, 40-49 and ≥50 years) and the number of that were present. These monthly values were then used to calculate the age-specific average rate over the study period.

This average (denoted $\overline{m}_a$) was then adjusted using the mortality rate among TB cases and the age-specific prevalence of TB in the final prevalence TB ($p_{TB,a}$) and mortality rate among TB cases before they start TB treatment (μ_{TB}) to obtain the average out-migration rate among individuals not on TB treatment $m_{tr-,a}$ using the following equation:

$$m_{tr-,a} = (\overline{m}_a - p_{TB}\mu_{TB})/(1-p_{TB}) \qquad \text{Equation 2.23}$$



Web Figure 7: Summary of the assumed monthly out-migration rates from each cluster for those who were on tuberculosis treatment or in different age groups.

2.7.4 Inmigration

A proportion of miners were assumed to have culture-positive TB on entry (see below), a proportion of whom were detected at the initial screening carried out when they joined the workforce (section 2.4.3) and were distributed among the compartments describing detected cases according to their relative size before the introduction of ART. The cases who were not detected were distributed into the compartments describing undetected cases according to their relative size before the introduction of ART. Similarly, non-diseased inmigrants were distributed between the latent, reinfected and recovered compartments according to their relative size before the introduction of ART.

The prevalence of TB among new mining employees is unknown. In the base case, it was assumed to equal the average age-specific prevalence measured in the final prevalence survey, after dividing by 0.98 to adjust for the effect of the intervention. This calculation implicitly assumed that the IPT intervention led to a 2% reduction in the TB prevalence. In sensitivity analyses, we explored the effect of alternative values for the prevalence (0%, 30% or 70% of that measured in the final prevalence survey, after adjusting for the effect of the

intervention). The HIV prevalence among new recruits without TB disease was assumed to be 30%, which is consistent with the average HIV prevalence in the general workforce(*16*).

Web Appendix 3: Estimating the unknown parameters in the model

3.1 Method for fitting the models

3.1.1 The expression for the loglikelihood deviance

For each assumption about the effect of IPT, the unknown parameters for the base case model were estimated by fitting model predictions of the measured incidence and the prevalence for the corresponding periods using maximum likelihood to the "observed" data (see below). The fitting was carried out using an algorithm based on the simplex method of Nelder and Mead(*31*), and used observed and model estimates for two age groups (<40 and ≥40 years). The expression for the loglikelihood deviance is taken as the contribution to the loglikelihood deviance of the saturated model and fitted model, summed for each cluster and age group. The expression for the loglikelihood for the incidence for each cluster for each age group is based on the Poisson distribution; that for the prevalence is based on the Binomial distribution. The expression for the loglikelihood deviance was as follows:

$$
\begin{aligned}
&-2\Bigg\{\sum_{j=1}^{16}\sum_{a=1}^{2} C_{j,a,PMP} - \hat{C}_{j,a,PMP} + C_{j,a,PMP}\ln(\hat{C}_{j,PMP}) - C_{j,a,PMP}\ln(C_{j,PMP}) \\
&+ K_{j,a,FPS}\ln(\hat{p}_{j,a,FPS}) + (N_{j,a} - K_{j,a,FPS})\ln(1-\hat{p}_{j,a,FPS}) - K_{j,FPS}\ln(p_{j,a,FPS}) \\
&-(N_{j,a} - K_{j,a,FPS})\ln(1-p_{j,a,FPS})\Bigg\}
\end{aligned}
$$

where $C_{j,a,PMP}$, $\hat{C}_{j,a,PMP}$, $N_{j,a}$, $K_{j,a,FPS}$, $p_{j,a,FPS}$ and $\hat{p}_{j,a,FPS}$ are defined in Web Table 8. Since the loglikelihood deviance expression uses model predictions of the incidence and prevalence for each cluster, it accounts for correlations between the incidence and the prevalence.

Web Table 8: Definitions of the statistics used in the fitting

Variable	Definition
$C_{j,a,PMP}$	The number of cases in age group *a* starting TB treatment during the primary measurement period in cluster *j* in the "observed" data.
$\hat{C}_{j,a,PMP}$	Model prediction of the number of cases in age group *a* starting TB treatment during the primary measurement period in cluster *j.* This was standardized to the population size used in the measured incidence in the observed data
$N_{j,a}$	The observed number of participants of age *a* in cluster *j* in the final prevalence survey
$K_{j,a,FPS}$	The number of cases of age *a* in cluster *j* with culture-positive TB in the final prevalence survey.
$p_{j,a,FPS}$	"Observed" proportion of individuals of age *a* in cluster *j* with culture-positive TB in the final prevalence survey.
$\hat{p}_{j,a,FPS}$	Model prediction of the proportion of individuals of age *a* in cluster *j* with culture-positive TB in the final prevalence survey

For the eight intervention clusters, the fitting procedure used the values for the numbers of cases starting TB treatment during the primary measurement period ($C_{j,a,PMP}$) and the prevalence of culture-positive TB during the final prevalence survey ($p_{j,a,FPS}$) that were seen in the trial.

For each simulated control cluster, the "observed" value for the numbers of cases starting TB treatment during the primary measurement period which was used in the fitting was taken to equal the value for the corresponding intervention cluster, after dividing it by the overall observed rate ratio (i.e. 1.02) describing the impact of the intervention on the measured incidence. Since control clusters in the trial were not perfectly matched to the intervention clusters, using cluster-level estimates of the impact of the intervention in this calculation would be inappropriate.

Likewise, the value used for the "observed" prevalence of culture-positive TB in the final prevalence survey ($p_{j,a,FPS}$) for the simulated controls which was used in the fitting was taken to equal the value for the corresponding intervention cluster, after dividing it by the overall observed prevalence ratio (i.e. 0.98) describing the impact of the intervention on the prevalence.

95% confidence intervals on the unknown parameters for the base case model were approximated by the 95% range of the best-fitting values obtained from fitting to 200 bootstrap datasets, which were generated so that they were consistent with the observed impact (see section 3.2). Similarly, for each assumption about IPT, the 95% range of the impact of the intervention on the TB incidence and prevalence was calculated from the best-fitting values obtained from fitting to the bootstrap samples. Each of the 200 bootstrap datasets was generated using a different value for the rate ratio, which was obtained by sampling from the distribution of plausible values for the rate ratio (see section 3.2).

The fitting used the measured incidence of definite and probable measured TB incidence in the fitting. Definite and probable TB cases largely comprised cases who had had microbiological confirmation of TB, although a small proportion of them would have been defined on the basis of clinical features. Since the model does not explicitly distinguish between cases who started TB treatment just on the basis of their symptoms from those who did so following firm microbiological confirmation. The measured incidence of definite and probable TB during the primary measurement period in the model was therefore approximated by the number of cases starting TB treatment during this time after multiplying it by the observed 1-proportion of all cases seen in the primary measurement period for the corresponding company that were defined as being possible (p_{poss}). The equation for the measured incidence in the model for a given cluster *j* used in the fitting is as follows:

$$\hat{C}_{j,a,PMP} = (1-p_{poss})\frac{N_{j,a,meas}}{\hat{N}_{j,a,meas}}\sum_{h}\sum_{t=T_{PMPstart}}^{T_{PMPfin}} C_{j,a,h}(t,0)$$

where $N_{j,a,meas}$ is the population denominator used in the observed measured incidence for cluster *j*, $\hat{N}_{j,a,meas}$ is the population denominator used in the measured incidence for the model for cluster *j*, $T_{PMPstart}$ and T_{PMPfin} are the start and end of the primary measurement period

respectively, and $C_{j,a,h}(t,0)$ is the model prediction of the number of cases, HIV status *h*, age group *a* starting TB treatment at time *t* in cluster *j*.

The prevalence of culture-positive TB in cluster *j* and age group *a* is given by the following equations:

$$\hat{p}_{j,a,FPS} = \frac{\sum_s \sum_h \left\{ \sum_{s_o=0}^{T_{o\max}} E_{j,n,s,a,h}(T_{FPS}, s_o) + E_{j,x,s,a,h}(T_{FPS}, s_o) + \sum_{s_f=0}^{T_{f\max}} F_{j,s,a,h}(T_{FPS}, s_f) \right\}}{\hat{N}_{j,a}(T_{FPS})}$$

where T_{FPS} is the time of the final prevalence survey; $E_{j,n,s,a,h}(T_{FPS}, s_o)$ and $E_{j,x,s,a,h}(T_{FPS}, s_o)$ are the numbers of undetected cases, smear status *s*, HIV status *h* and age group *a* in cluster *j* experiencing disease through endogenous reactivation or through exogenous reinfection respectively, at time T_{FPS} and time s_o since onset; $F_{j,s,a,h}(T_{FPS}, s_f)$ is the number of cases, smear status *s*, HIV status *h* and age group *a* in cluster *j* who have been detected for duration s_f and have not yet started TB treatment at time T_{FPS}; $\hat{N}_{j,a}(T_{FPS})$ is the number of individuals of age *a* in cluster *j* at time T_{FPS}.

3.1.2 Constraints on the parameter values

When fitting the model, the parameters were constrained for biological plausibility as follows:

1. Rates of disease onset (either through reactivation or following reinfection) for HIV negative miners without radiologically-confirmed silicosis aged <40 years were less than those for similar miners aged ≥40 years.
2. Reactivation rate among HIV negative miners without radiologically-confirmed silicosis was less than or equal to 13% of the rate of disease onset during the first year following reinfection. This constraint follows from historical data in which the rate of disease onset appeared to decrease with time since initial infection(*9*). The

reduction in the rate of disease onset with time since reinfection is unknown; however, the assumed relationship is plausible.

3. Rate of disease onset during the first year after reinfection among HIV negative miners without radiologically confirmed silicosis was less than 11%/year. The estimates for Western populations have typically been about 5%/year(*10-11*).
4. Reactivation rate among HIV negative miners was less than 1.4%/year, which is consistent with constraints 2 and 3. Estimates for Western populations range between 0.03%/yr for white males in England and Wales or The Netherlands and 0.5% for elderly males in Hong Kong(*10-11*).
5. The proportion of infections that are cured by 6 months of IPT for HIV-negative individuals was at least as large as that for HIV-positive miners.

To ensure that the model only used positive values for parameters, we only passed the positive value of parameters to the function calculating the loglikelihood deviance. We imposed constraints by allowing the fitting algorithm to select any value for a parameter, but then reset it to be just within the boundary (e.g. 0.9999 if the upper limit was 1.0) if it was outside the permissible range. To increase the probability that the values selected by the fitting routine were globally optimum, we started the fitting process for 20 different starting values. The starting values were selected to span the range of plausible parameter values. All (apart from one or two) of the starting points converged to the same value. The ones that did not converge were associated with a very poor fit to the data. In addition, the implemented Nelder-Mead algorithm includes a local-restart procedure where following convergence, a new simplex is initialised from the local optima and the search repeated. This restart procedure was repeated 4 times for each initial starting value.

Since the rates of disease onset change with the fitting process, the prevalence of infectious individuals and therefore the average annual risk of infection also changes with the outcome of the fitting. The fitting process was therefore repeated for multiple values of the effective

contact rate, and the value for the effective contact rate used in the base case model subsequently was the one which led to an average annual risk of infection averaged across the clusters before the introduction of IPT which best matched the intended value.

3.2 Generating the bootstrap datasets

The bootstrap datasets which were used to calculate approximate 95% confidence intervals of the estimated parameters and the 95% range of the best-fitting impact for the 3 IPT assumptions (see section 3.1.1) were generated so that they were consistent with the impact observed in the overall incidence of definite and probable TB and prevalence during the intervention, i.e. incidence rate ratio: 1.04 (95% CI: 0.73-1.48), prevalence ratio: 0.98 (95% CI: 0.65-1.48).

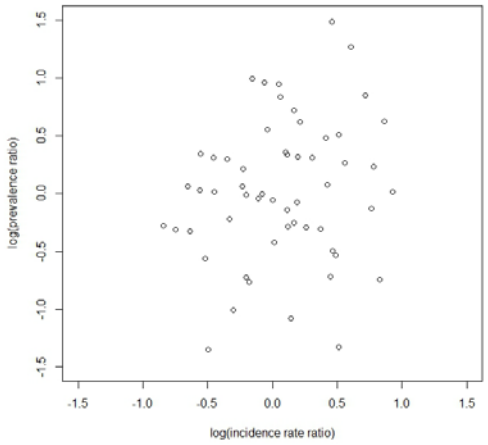
In all the bootstrap datasets, the measured incidence of definite and probable TB and the prevalence for each of the intervention clusters during the corresponding periods were taken to be identical to those observed during the study.

The value for the incidence for the simulated control for each intervention cluster was calculated by dividing the observed value for the corresponding intervention cluster by a random value for the overall average incidence rate ratio, which was sampled to be consistent with the 95% CI of the observed incidence rate ratio. The value for the prevalence for the simulated control for each intervention cluster was calculated similarly, i.e. by dividing the observed prevalence for the corresponding intervention cluster by a random value for the overall average prevalence ratio which was consistent with the 95% CI of the observed prevalence ratio.

Since, in reality, the impact on the incidence and prevalence in a given cluster was likely to be correlated, the method for sampling the incidence rate ratio and prevalence ratio used in these calculations took account of this correlation (see below).

3.2.1 The correlation between the incidence rate ratio and prevalence ratio

Since the control clusters in the Thibela trial were not perfectly matched to a given intervention cluster, the correlation, ρ, was calculated using all possible pairs of intervention and control clusters (8 intervention and 7 controls, 56 possible pairs). For each pair the log(prevalence ratio) and log(incidence rate ratio) are calculated as the difference between the log(prevalence) or log(incidence) in the intervention and control cluster. The correlation between the log(prevalence ratio) and log(incidence rate ratio) was then calculated. The values are shown in Web Figure 8 and the resulting correlation is ρ=0.18.



Web Figure 8: Scatterplot of the log(prevalence ratio) against the log(incidence rate ratio) calculated for each of the 56 possible pairings of incidence and control clusters.

3.2.2 Sampling prevalence and incidence rate ratios

When sampling values for the prevalence and incidence rate ratios, it was assumed that the overall log(prevalence ratio) and log(incidence rate ratio) followed normal distributions with correlation ρ. Using this assumption pairs of values for the prevalence and rate ratio were generated using the following algorithm:

1. Pick values *u* and *v* from independent standard normal distributions

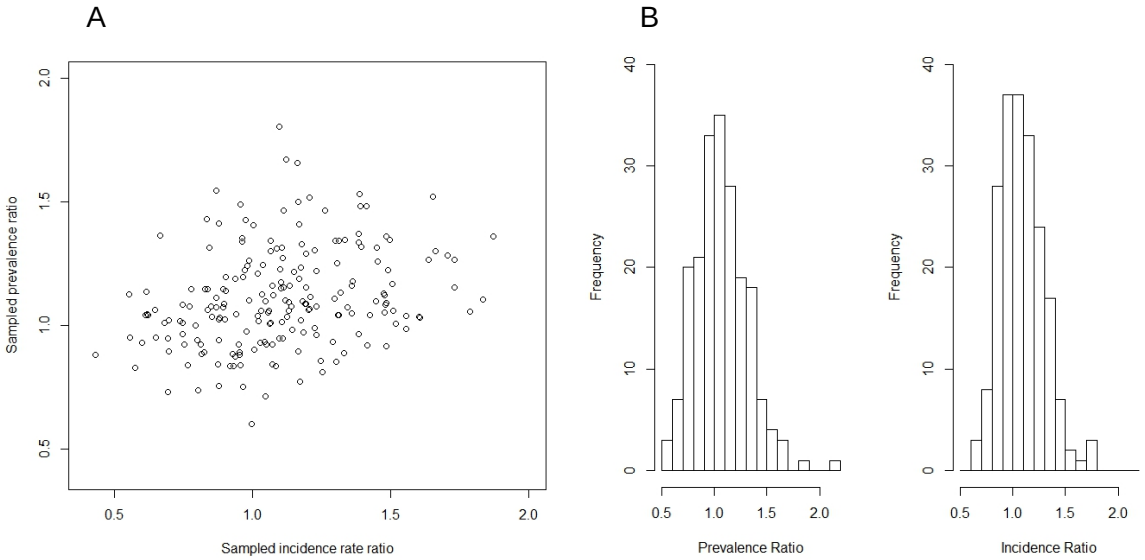
2. Calculate the log(prevalence ratio) = $s_{prevalence}u + m_{prevalence}$
3. Calculate log(incidence rate ratio) = $s_{incidence}\left(\rho u + v\sqrt{1-\rho^2}\right) + m_{incidence}$

   where:

   $s_{prevalence}$ and $s_{incidence}$ are the standard deviations of the log(prevalence ratio) and log(incidence rate ratio) respectively, $m_{prevalence}$ and $m_{incidence}$ are the means of the log(prevalence ratio) and log(incidence rate ratio) respectively and ρ is the correlation between log(prevalence ratio) and log(incidence rate ratio)
4. Take the exponential of the log values calculated in steps 2 and 3 to obtain the prevalence and incidence rate ratios

The mean log(prevalence ratio) and log(incidence rate ratio) (and standard deviation) were estimated using linear regression of the log(prevalence) and log(incidence) on stratum and arm(*27*).

Using this approach, 200 samples for the prevalence ratio and incidence rate ratio were generated. Web Figure 9A shows a scatter plot of the sampled prevalence and incidence rate ratios. The distributions of these values are shown in Web Figure 9B.



Web Figure 9A. Scatterplot of the 200 sampled prevalence ratios and incidence rate ratios. B**.** Distribution of the 200 sampled prevalence and incidence ratios.

Web Appendix 4: Modelling interventions which might help to control TB in the mines

4.1 Reducing initial loss to follow-up and time to treatment (without Xpert MTB/RIF)

For objectives 2 and 3 (the maximum achievable impact in Thibela TB and what might control TB in goldmines respectively), the initial loss to follow-up was assumed to be 5%. For objective 2 and objective 3 in the absence of Xpert MTB/RIF, the rates at which TB cases started treatment were recalculated using the equations below (adapted from those in section 2.5) so that 90% of smear-negative cases started treatment within 5 weeks, and 90% of smear-positive cases starting treatment within 1 month:

$$u(s_f + 1) = u(s_f) - \tau(s_f)u(s_f)$$
$$u(26) = i_D$$
$$u(T_{90}) = 0.1$$

where $\tau(s_f)$ is the rate at which cases start TB treatment in week s_f after detection (assumed to be constant in each month and differs between the first and subsequent months); $u(s_f)$ is the estimated proportion of those detected who are still untreated s_f weeks after detection; i_D is the initial loss to follow-up; T_{90} is the time by which 90% of those detected are assumed to have started TB treatment.

These equations lead to the values for the rate at which detected cases start TB treatment and the average time to treatment among those who are treated shown in Web Table 9.

Web Table 9: Assumed values for the average time to treatment among treated miners

	Month 1	**Months 2-6**	**Average time to treatment (weeks)**
Smear-positive	43.8%/week	3.1%/week	2.4
Smear-negative	37.9%/week	4.8%/week	3.1

The average time to treatment is calculated using the following equation:

$$T_{TR_{AVE}} = \frac{\sum_{s_f=0}^{26} s_f \tau(s_f) u(s_f)}{\sum_{s_f=0}^{26} \tau(s_f) u(s_f)}$$

4.2 Provision of continuous IPT

The equations used when implementing provision of continuous IPT were identical to those presented in section 1.2, except for the inclusion of additional compartments for those in the latent, reinfected and recovered compartment who were on IPT continuously (denoted using the symbols $L_{z_\infty}(t)$, $R_{z_\infty}(t, s_r, s_z)$ and $V_{z_\infty}(t)$ for individuals at time *t*, where s_r and s_z are the times since reinfection and the start of IPT respectively).

Provision of IPT continuously in the model was implemented by first providing 9 months of IPT for all individuals, with coverage & retention equalling those in the best-performing cluster (cluster 7) in Thibela TB. Any individuals who were still on IPT at the end of the 9 month period (approximately 50% of each cluster) were then transferred into the corresponding compartment for those on continuous IPT. To ensure that 50% of the population was on continuous IPT subsequently, 50% of new mining employees without tuberculosis disease were allocated to the corresponding compartment for those on continuous IPT.

To simplify the equations, individuals in the compartments relating to continuous IPT are not allowed to cure their infections, which is consistent with findings in objective 1 that IPT cured only a small proportion of latent infections.

The equations relating to $L_{z_\infty}(t)$, $R_{z_\infty}(t, s_r)$ and $V_{z_\infty}(t)$ are provided below as are any amendments to Equation 1.1-Equation 1.16. For simplicity, any equations in the Thibela model which remain unchanged with the introduction of continuous IPT are not reproduced.

Individuals with latent infection

$$L_{z-,a,h}(t+\delta t) = L_{z-,a,h}(t)(1-i_{z+,a}(t))(1-d_{n,a,h,z-}k_{si,a,h} - \lambda(t) - m_{tr-,a}) + M_{in,L,a,h}(t)(1-p_{z_\infty}) + R_{z-,a,h}(t,T_R)$$ Equation 1.1a'

$$L_{z_\infty,a,h}(t+\delta t) = L_{z_\infty,a,h}(t) - L_{z_\infty,a,h}(t)(d_{n,a,h,z+}k_{si,a,h} + (1-\pi_{r,z+,h})\lambda(t) + m_{tr-,a}) + M_{in,L,a,h}(t)p_{z_\infty} + L_{z+,a,h}(t,T_{z_{\max}}) + R_{z_\infty,a,h}(t,T_R) + R_{z+a,h}(t,T_R,T_{z_{\max}})$$ Equation 4.1

Miners who have completed IPT and have not been reinfected in the preceding two years

$$P_{e+,a,h}(t+\delta t) = P_{e+,a,h}(t) - (\lambda(t) + m_{tr-,a})P_{e+,a,h}(t)$$ Equation 1.2a'

$$P_{e-,a,h}(t+\delta t) = P_{e-,a,h}(t) - P_{e-,a,h}(t)(d_{n,z-,a,h}k_{si,a,h} + \lambda(t) + m_{tr-,a}) + R_{z_p}(t,T_R)$$ Equation 1.2b'

Reinfected miners who are not on IPT

$$R_{z-,a,h}(t+\delta t, s_r + \delta s_r) = (1-i_{z+}(t))R_{z-,a,h}(t,s_r)(1-d_{x,z-,a,h}(s_r)k_{si,a,h} - m_{tr-,a}) + M_{in,R,a,h}(t,s_r)(1-p_{z_\infty})$$ Equation 1.3b'

Reinfected miners who are on continuous IPT

$$R_{z_\infty,a,h}(t+\delta t,0) = (1-\pi_{r,z+,h})\lambda(t)(L_{z_\infty,a,h}(t) + V_{z_\infty,a,h}(t))(1-d_{x,z+,a,h,}(s_r)k_{si,a,h})$$ Equation 4.2a

$$R_{z_\infty,a,h}(t+\delta t, s_r + \delta s_r) = R_{z_\infty,a,h}(t,s_r) - R_{z_\infty,a,h}(t,s_r)(d_{x,z+,a,h,}(s_r)k_{si,a,h} + m_{tr-,a}) + M_{in,R,a,h}(t,s_r)p_{z_\infty}$$ Equation 4.2b

Cases experiencing disease through reactivation, who have not yet been detected:

$$E_{n,s-,a,h}(t+\delta t,0) = L_{z-,a,h}(t)d_{n,z-,a,h}k_{si,a,h} + (L_{z+,a,h}(t) + L_{z_\infty,a,h}(t))d_{n,z+,a,h}k_{si,a,h}$$ Equation 1.6a'

Cases experiencing disease because of reinfection, who have not yet been detected:

$$E_{x,s-,a,h}(t+\delta t,0) = \sum_{s_r=0}^{T_R} d_{x,z-,a,h}(s_r)k_{si,a,h}(R_{z-,a,h}(t,s_r)+R_{z_p,a,h}(t,s_r)) + \sum_{s_r=0}^{T_R}\sum_{s_z=0}^{T_{z_{max}}} R_{z+,a,h}(t,s_r,s_z)d_{x,z+,a,h}(s_r)k_{si,a,h} + \sum_{s_r=0}^{T_R} R_{z_\infty,a,h}(t,s_r)d_{x,z+,a,h}(s_r)k_{si,a,h}$$ Equation 1.7a'

Miners who have recovered from TB disease:

$$V_{z-,a,h}(t+\delta t) = V_{z-,a,h}(t) - V_{z-,a,h}(t)(i_{z+,a}(t)+\lambda(t)+m_{tr-,a}) + M_{in,V,a,h}(t)(1-p_{z_\infty}) + C(T_{\tau_{max}})$$ Equation 1.10a'

$$V_{z_\infty,a,h}(t+\delta t) = V_{z_\infty,a,h}(t) - V_{z_\infty,a,h}(t)(\lambda(t)(1-\pi_{r,z+,h})+m_{tr-,a}) + M_{in,V,a,h}(t)p_{z_\infty} + V_{z+,a,h}(t,T_{z_{max}})$$ Equation 4.3

Transitions at the end of each time step:

$$P_{e+,a,h}(t+\delta t) = P_{e+,a,h}(t) + R_{z+,a,h}(t,T_{z_{min}},T_{z_{max}})p_{z_c,h} + L_{z+,a,h}(t,T_{z_{min}})i_{z-}(T_{z_{min}})p_{z_c,h} + \sum_{s_z=0}^{T_{z_{max}}-1} V_{z+,a,h}(t,s_z)i_{z-}(s_z) + \sum_{s_r\geq T_{z_{min}}} R_{z+,a,h}(t,s_r,T_{z_{min}})i_{z-}(T_{z_{min}})p_{z_c,h} + \sum_{s_z\geq T_{z_{min}}} R_{z+,a,h}(t,T_{z_{min}},s_z)i_{z-}(s_z)p_{z_c,h}$$ Equation 1.11a'

$$P_{e-,a,h}(t+\delta t) = P_{e-,a,h}(t) + L_{z+,a,h}(t,T_{z_{min}})i_{z-}(T_{z_{min}})(1-p_{z_c,h}) + \sum_{\substack{s_z=0,\\ s_z\neq T_{z_{min}}}}^{T_{z_{max}}-1} L_{z+,a,h}(t,s_z)i_{z-}(s_z) + R_{z+a,h}(t,T_R,T_{z_{max}})$$ Equation 1.11b'

$$V_{z_\infty}(t+\delta t) = V_{z_\infty}(t) + V_{z+}(t,T_{z_{max}}) + R_{z+}(t,T_{z_{min}},T_{z_{max}})p_{z_c}$$ Equation 4.4

$$L_{z_\infty}(t+\delta t) = L_{z_\infty}(t) + L_{z+}(t,T_{z_{max}}) + R_{z+}(t,T_R,T_{z_{max}})$$ Equation 4.5

$$R_{z_\infty,a,h}(t+\delta t,s_r+\delta s_r) = R_{z_\infty,a,h}(t,s_r) + R_{z+,a,h}(t,s_r,T_{z_{max}}) \quad s_r < T_{z_{min}}$$ Equation 4.6a

$$R_{z_\infty}(t+\delta t,s_r+\delta s_r) = R_{z_\infty}(t,s_r) + R_{z+}(t,T_{z_{min}},T_{z_{max}})(1-p_{z_c}) \quad s_r = T_{z_{min}}$$ Equation 4.6b

$$R_{z_\infty}(t+\delta t,s_r+\delta s_r) = R_{z_\infty}(t,s_r) + R_{z+,a,h}(t,s_r,T_{z_{max}}) \quad s_r > T_{z_{min}}$$ Equation 4.6c

4.3 Scale up of ART

The equations used in the model to explore the effect of ART scale up to different groups of HIV-positive people were identical to those presented in section 1.2, except for the scaling

factor $k_{z-,ART,h}(t)$ (which reflects the factor by which the average rate of disease onset among HIV-positive miners without radiologically-confirmed silicosis differs from that for HIV-negative miners without radiologically-confirmed silicosis – see section 2.2.1) was adapted to be as follows:

$$k_{z-,ART,h}(t) = \sum_i p_{h,c_i-c_{i+1}} g_{ART,c_i-c_{i+1}}(t)\rho_{h,c_i-c_{i+1}}(1-\pi_{d,z-,ART+}) + p_{h,c_i-c_{i+1}}(1-g_{ART,c_i-c_{i+1}}(t))\rho_{h,c_i-c_{i+1}}$$

where:

$p_{h,c_i-c_{i+1}}$ is the proportion of HIV-positive people in CD4 stratum c_i-c_{i+1}. The CD4+ cell count strata used were <200, 200-<350, 350-<500, ≥500 cells/μL and $p_{h,c_i-c_{i+1}}$ was taken to equal 0.25 in each of these strata (see section 2.3.2.1).

$g_{ART,c_i-c_{i+1}}(t)$ is the ART coverage among those in CD4 stratum c_i-c_{i+1},

$\rho_{h,c_i-c_{i+1}}$ is the factor by which the rate of disease onset among those with a CD4+ cell count in the range c_i-c_{i+1} cells/μL differs from that among HIV-negative individuals, assumed to equal 17, 9.7, 5.8, 2.8 for those with CD4 counts in the range <200, 200-<350, 350-<500 and ≥500 cells/μL;

$\pi_{z-,ART+}$ is the protection provided by ART for those who are not on IPT, assumed to equal 65% (see section 2.3.3).

The corresponding scaling factor applied to obtain the rate of disease onset among HIV-positive individuals (without silicosis) who are both on IPT and on ART ($k_{z+,ART,h}$) was calculated as follows:

$$\begin{aligned} k_{z+,ART,h}(t) = \sum_i & p_{h,c_i-c_{i+1}} g_{ART,c_i-c_{i+1}}(t)\rho_{h,c_i-c_{i+1}}(1-\pi_{d,z+,ART+}) \\ & + p_{h,c_i-c_{i+1}}(1-g_{ART,c_i-c_{i+1}})(t)\rho_{h,c_i-c_{i+1}}(1-\pi_{d,z+,ART-}) \end{aligned}$$

where $\pi_{z-,ART+}$ is the protection provided by ART for those who are on IPT, assumed to equal 82.5% (see section 2.3.3).

4.4 Improved diagnosis of cases using Xpert MTB/RIF

4.4.1 Overview

The introduction of Xpert MTB/RIF into the health services is likely to lead to changes in the proportion of miners with TB that are detected upon presentation either when they join the workforce or when they present passively or at the routine medical exam and to reductions in initial loss to follow-up. In these analyses, we explore the effect of two assumptions relating to Xpert MTB/RIF namely that :

a. Scenario 1: Xray is still used to screen miners presenting at routine medical exam or when joining the workforce, but Xpert MTB/RIF is used to diagnose people with suspected TB both in the routine medical exam and on passive presentation.
b. Scenario 2: Xpert MTB/RIF is used both to screen and diagnose cases in the routine medical exam and for people who present passively.

4.4.2 Sensitivity

The sensitivity of Xpert MTB/RIF is assumed to be 55% and 97% for smear-negative and smear-positive TB respectively, based on findings from a sub-study carried out during the final prevalence survey in Thibela TB(*32*). The sensitivity of Xpert, compared to MGIT as the gold standard was 55.2% (95% CI: 47-63%) and 97% (95% CI: 84-100%) for smear-negative and smear-positive TB respectively.

These estimates are consistent with findings from other population-based surveys in high HIV-prevalence settings in which only one cartridge was used when evaluating Xpert MTB/RIF (*33-34*). For example, one study(*35*), which considered cases enrolling at ART clinics who did not have a current diagnosis of TB or a previous history of taking ART, found

a sensitivity of 43%. Another study, who considered individuals with suspected TB attending primary care clinics in Cape Town, found a sensitivity of 55% (95% CI: 35-73%)(*34*).

Literature reviews have identified values for the sensitivity which exceed 50%(*35*). However, these estimates come from studies considering selective populations under conditions which are unlikely to be achieved by the mine health services, e.g. individuals who probably had a very high bacillary load and/or using two or more cartridges when diagnosing cases or exerting extra efforts when extracting sputum from individuals.

4.4.3 TB treatment

4.4.3.1 Initial loss to follow-up

In these analyses, the effect of Xpert MTB/RIF is explored after the health systems have been improved to reduce initial loss to follow-up to 5%.

In the absence of such improvements in the health systems, the introduction of Xpert/MTB RIF would probably not affect the initial loss to follow-up among smear-positive cases, since the time until the Xpert MTB/RIF results return from the lab is comparable to that relating to smear microscopy. For smear-negative TB, the introduction of Xpert/MTB RIF would probably lead to reductions in the initial loss to follow-up so that it equals the corresponding value for smear-positive TB, since the time until Xpert/MTB RIF lab results are received is identical for smear-negative and smear-positive cases. This is consistent with findings from one multicentre study considering urban health centres(*36*), which suggested that the proportion of cases who were not treated decreased from 39·3% (95% CI 32·6–46·6) to 14·7% (9·9–21·2) after Xpert MTB/RIF was introduced.

4.4.3.2 Time to TB treatment

The introduction of Xpert MTB/RIF is assumed to lead to a reduction in the time between detection and TB treatment among the cases who eventually start TB treatment to an average period of 2 weeks, irrespective of smear status. This is consistent with findings (*36*)

in which the median time-to-treatment for smear-negative tuberculosis reduced from 56 to 5 days as a result of introducing Xpert MTB/RIF.

Applying the equations in section 4.1 (without the constraint for the time by which 90% of miners have started treatment after detection) to reproduce the assumed initial loss to follow-up (5%, occurring once the health systems have been improved) leads to values for the rate at which detected cases start treatment of 50.4%/week in the first month after detection and 0.8%/week subsequently.

4.4.4 Case detection rates

4.4.4.1 Routine medical exam

4.4.4.1.1 <u>The rate at which cases are detected through the routine medical exam</u>

The rate at which smear-positive cases are detected through the routine medical exam was calculated using the same equation as that used considering the situation before the introduction of Xpert MTB/RIF (Equation 2.14 - reproduced below), assuming that the time required to call up the workforce remains unaffected:

$$r_{x1,s,y} = 1-(1-0.95p_{x1,s,y})^{1/64} \qquad \text{Equation 2.14}$$

where $p_{x1,s,y}$ is the proportion of cases with smear status *s* attending the routine medical exam screen that are detected following the introduction Xpert MTB/RIF. As shown below (section 4.4.4.1.2), this proportion remains unchanged for smear-positive cases after the introduction of Xpert MTB/RIF in scenario 1, equalling 0.25 for both companies. The corresponding proportion considering smear-negatives is 0.137 and 0.075 for companies A and B/C respectively (section 4.4.4.1.2). For smear-positive cases, Equation 2.14 leads to values of $r_{x1,s,y}$ of 0.423%/week for all companies and to values of 0.217% and 0.116%/week for smear-negative cases for companies A and B/C respectively.

For Xpert scenario 2, the proportion of cases attending the routine medical exam that are detected equals the sensitivity of Xpert MTB/RIF for detecting smear-negative and smear-positive TB, i.e. 55% and 97% respectively (section 4.4.2), which is identical for all companies. Substituting these values into Equation 2.14 leads to the following values for the rates at which smear-positive and smear-negative cases are detected through the routine medical exam (which are increased compared to the values before the introduction of Xpert MTB/RIF):

$$r_{x2,s+,y} = 1-(1-0.95\times0.97)^{1/64} = 0.0390 \text{ per week} \qquad \text{Equation 4.7}$$

$$r_{x2,s-,y} = 1-(1-0.95\times0.55)^{1/64} = 0.0115 \text{ per week} \qquad \text{Equation 4.8}$$

4.4.4.1.2 Proportion of cases attending the routine medical exam that are detected - scenario 1

The proportion of smear-positive cases attending the routine medical exam that are detected following the introduction of Xpert MTB/RIF in scenario 1, $p_{x1,s+,y}$ is calculated using the same equation as that used in the absence of Xpert MTB/RIF (Equation 2.15, reproduced for convenience here):

$p_{x1,s+,y}$ =
(sensitivity of X-ray for detecting smear-positive TB)
×
(proportion of cases that are investigated further following the X-ray)
× Equation 2.15
(sensitivity of the method used in further investigations)

For all companies, it is assumed that the proportion of TB suspects, based on X-ray, that are investigated further for TB remains unchanged in this scenario and equals 65%.

For company A, the isolates of all cases who return for subsequent investigation are cultured even without the introduction of Xpert MTB/RIF. Since Xpert MTB/RIF has a lower

sensitivity than culture, it is assumed that its introduction does not affect the proportion of cases that are detected in the routine medical exam, and these remain as follows:

smear-negative: $p_{x1,s-,y} = 0.21 \times 0.65 \times 1.00 = 0.1365$

smear-positive: $p_{x1,s+,y} = 0.385 \times 0.65 \times 1.00 = 0.2505$

For companies B and C, it is assumed Xpert MTB/RIF identifies 55% (i.e. the sensitivity of Xpert MTB/RIF) of smear-negative TB cases who return for subsequent investigation, as compared with 10% before the introduction of Xpert MTB/RIF. However, we assume that introducing Xpert MTB/RIF does not affect the proportion of smear-positive cases that are detected in the routine medical exam, since sputum isolates will probably continue to be examined with smear microscopy since Xpert MTB/RIF identifies <100% (i.e. 97%) of smear-positive cases. The corresponding proportions are as follows for companies B and C:

smear-negative: $p_{x1,s-,y} = 0.21 \times 0.65 \times 0.55 = 0.0751$

smear-positive: $p_{x1,s+,y} = 0.385 \times 0.65 \times 1.00 = 0.2505$

4.4.4.2 Rates at which cases are detected following passive presentation

4.4.4.2.1 Smear-positive cases

For both Xpert scenarios, it is assumed that the rates at which smear-positive cases are detected through passive presentation remain unchanged following the introduction of Xpert MTB/RIF, given that all companies are identifying smear-positive cases at their optimum levels before it is introduced, by using smear microscopy (see section 2.4.2.1).

4.4.4.2.2 Smear-negative cases

The rates at which smear-negative cases are detected through passive presentation are assumed to be identical for both Xpert MTB/RIF scenarios, but differ between the mining companies. For simplicity, it is assumed that cases are as likely to present passively to the

health services for diagnosis after the introduction of Xpert MTB/RIF as they were before its introduction.

For company A, isolates from all cases who present passively are cultured even in the absence of Xpert MTB/RIF. The introduction of Xpert MTB/RIF, in addition to culture to diagnose these cases, will probably result in cases being detected on average 6 weeks earlier than in the absence of Xpert MTB/RIF. The rate at which smear-negative cases with HIV status *h* are detected through passive presentation in company A is calculated as follows:

$$r_{x1,s\text{-},h,p} = 1/(D_{s\text{-},h,p} - 6) \qquad \text{Equation 4.9}$$

where $D_{s\text{-},h+}$ is the average time to detection through passive presentation of smear-negative cases (in weeks) with HIV status *h*, calculated as 1/(rate at which cases are detected through passive presentation) or $1/r_{f,s\text{-},h,p}$.

Substituting the corresponding values for $r_{f,s\text{-},h,p}$ (see Equation 2.18 and Equation 2.19) into Equation 4.9 leads to the following values for the rate at which smear-negative cases are detected through passive presentation following the introduction of Xpert MTB/RIF:

HIV-positives: $r_{x1,s\text{-},h+,p} = 1/((1/0.0124) - 6) = 0.0134$ per week Equation 4.10

HIV-negatives: $r_{x1,s\text{-},h\text{-},p} = 1/((1/0.0041) - 6) = 0.0042$ per week Equation 4.11

These values are similar to the corresponding values before the introduction of Xpert MTB/RIF (0.0124 per week and 0.0041 per week for HIV-positives and HIV-negatives respectively – see Web Table 10).

4.4.4.2.2.1 Companies B and C

In contrast with company A, where cases are detected using culture, a proportion of smear-negative cases in companies B and C are probably referred for treatment on the basis of clinical symptoms. The introduction of Xpert MTB/RIF is therefore unlikely to change the rate at which these cases are detected and put onto treatment.

We therefore assume that the rates at which smear-negative cases in companies B and C are detected through passive presentation following the introduction of Xpert MTB/RIF equal the values assumed from before the introduction of Xpert MTB/RIF, i.e. 1.44%/week and 0.61%/week for HIV-positive and HIV-negative TB cases respectively (see section 2.4.2.3).

the overall rates at which smear-negative and smear-positive cases are detected (calculated as the sum of the rates at which cases are detected through passive presentation and routine medical exam) are similar for all companies, apart from that for HIV-negative smear-negative cases (Web Table 10). For these miners, the rates at which they are detected following Xpert MTB/RIF is introduced are slightly higher for companies B and C than for company A.

4.4.4.3 Initial screen on joining the workforce

It is assumed that the procedure for screening miners when they join the workforce is identical to that used for screening miners at the routine medical exam. As is the case before the introduction of Xpert MTB/RIF, the proportion of cases that are detected when joining the mines is then identical to the proportion detected at the routine medical exam and differs between companies and according to smear status (see Web Table 10).

Web Table 10: Summary of the assumptions relating to case detection rates, before and after the introduction of Xpert MTB/RIF. The shaded cells show the values which have changed as a result of the introduction of Xpert.

Definition	**Company**		**Pre-Xpert**		**Xpert scenario 1**		**Xpert scenario 2**	
			Smear-	**Smear+**	**Smear-**	**Smear+**	**Smear-**	**Smear+**
Proportion of cases that are detected when attending the routine medical exam	A	Both	0.1365	0.2505	0.1365	0.2505	0.55	0.97
	B/C	Both	0.0137	0.2505	0.0751	0.2505	0.55	0.97
Rate at which cases are detected through the routine medical exam(%/week)	A	Both	0.217	0.423	0.217	0.423	1.148	3.898
	B/C	Both	0.020	0.423	0.116	0.423	1.148	3.898
Rate at which cases are detected through passive presentation (%/week)	A	HIV+	1.244	12.95	1.34	12.95	1.34	12.95
		HIV-	0.4094	1.55	0.4197	1.55	0.4197	1.55
	B/C	HIV+	1.440	12.95	1.440	12.95	1.440	12.95
		HIV-	0.6058	1.55	0.6058	1.55	0.6058	1.55
Overall rates at which cases are detected (%/week)	A	HIV+	1.46	13.40	1.566	13.40	2.49	16.84
		HIV-	0.626	1.97	0.6365	1.97	1.57	5.45
	B/C	HIV+	1.46	13.40	1.556	13.40	2.59	16.84
		HIV-	0.626	1.97	0.7231	1.97	1.75	5.45

4.5 Reducing the prevalence of silicosis

The proportion of cases attributable to HIV and silicosis in the model was estimated using an approach which has been used elsewhere (*37*). Specifically, the model was first run with the default values for the relative risk of TB for silicotics and HIV-positive individuals, compared to non-silicotics or HIV-negative individuals respectively (Web Table 11). It was then re-run to calculate the incidence from a given time T in the absence of silicosis and HIV, by adjusting the relevant risk factors accordingly (i.e. the relative risk of TB for silicotics and HIV-positive individuals, compared to non-silicotics or HIV-negative individuals set to 1.0). For each model run, the total number of new TB cases in the year the co-factor was removed was calculated. A one year window was selected to minimise the indirect effect on the number of new cases which would result from the reduction in prevalence caused by removing the co-factors. The proportion attributable fraction (PAF) was then calculated as follows:

$$PAF_R = 100 \times \left(1 - \frac{C_{-R}}{C_{+R}}\right)$$

where C_{+R} is the number of cases with the default risk factors and C_{-R} is the number of cases calculated after the corresponding relative risk has been set to 1.0. When calculating the PAF for HIV, the relative risk for HIV-positive silicotics was set equal to the value for HIV-negative individuals with silicosis ($\rho_{si,h+} = \rho_{si,h-}$) to adjust for the higher risk of TB in due to silicotics in HIV infected individuals compared to HIV negative members of the population(*7*).

PAF values were calculated for each cluster individually and across all clusters combined. The PAF for both silicosis and HIV were calculated at two time points in the model; three years prior to the introduction of ART and at the time of the intervention. This allowed the impact of ART on the PAF for HIV to be evaluated. To ensure the intervention did not affect the calculations they were performed using the predicted numbers of cases in the simulated control clusters.

The sensitivity of the PAF to the HIV prevalence, ARI and prevalence of culture-positive TB among new mining employees was explored by calculating the PAF for the following values for these parameters, using the best fitting disease rates obtained by fitting the model to the observed data (see Web Appendix 3).

Average ARI before the introduction of IPT: 10%, 20% (base case), 30%/year

HIV prevalence: 20%, 30% (base case), 35%, 40%

Prevalence of culture-positive TB among new mining employees: assumed to differ by the following factors from that seen in the final prevalence survey – 0%, 33%, 66%, 100%

Web Table 11: Risk factors changed when calculating the PAF for silicosis and HIV

Parameter		**Symbol**	**Defaultvalue**
Relative risk of TB in HIV+ compared to those HIV-	>200 cells/mL	$\rho_{h,>200}$	5.9
	<200 cells/mL	$\rho_{h,<200}$	17.01
Protection provided by ART against TB		$\pi_{z-,ART+}$	0.65
Relative risk of TB in radiologically-confirmed silicotics, compared miners without radiologically-confirmed silicosis	HIV-	$\rho_{si,h-}$	2.6
	HIV+	$\rho_{si,h+}$	4.1

Web Appendix 5: Additional Results

5.1 Objective 1: Understanding the lack of population-level impact

Table 5.1 shows the best fitting parameter values, predicted impacts and goodness of fit statistic (deviance) for each IPT model for different assumptions about the ARI, HIV prevalence and prevalence of disease among new mining employees.

The range of impacts predicted by the model is narrower than the 95% CI calculated from the trial observations. At one extreme, the trial data includes prevalence and incidence rate ratios greater than one (consistent with a negative trial impact). However the model is not able to predict a negative impact (which would require IPT to increase the risk of tuberculosis). For higher observed effects the predicted model impacts are limited by fitting to incidence and prevalence data simultaneously. Due to the low correlation between the prevalence ratio and incidence rate ratio observed in the trial many of the bootstrap samples combine positive and negative impacts in the two measured outcomes.

Web Table 12: Best-fitting values obtained for each of the IPT assumptions. The 95% confidence intervals obtained by bootstrapping are indicated in red; those obtained by profile likelihood are in black font. Differences in the intervals between the two approach result from the constraints placed upon the parameter values in the fitting process, which ensure that the proportions cured and protected by IPT are lower in HIV positive individuals than in HIV negative individuals. Thus when the 95% CI are generated by bootstrapping for the protection provided by IPT for HIV-positive individuals, it is constrained to be less than that provided for HIV-negative individuals, whereas this constraint is not in place when 95% CI are calculated by profile likelihood.

ARI (%)	HIV (%)	Mig (%)	Disease rates (HIV negative) (%/year)				% cured		%protected		Impacts (%)		% due to reinfn	Deviance
			Reactivation (d_n)		Exogenous* (d_x)									
			Age<40	Age≥40	Age<40	Age≥40	HIV-	HIV+	HIV-	HIV+	Incidence	Prevalence		
IPT assumption #1 (100% cure, 100% protection)														
10	30	100	1.3 (0.8-1.4)	1.4 (1.3-1.9)	11.0 (9.2-11.6)	11.0 (10.5-11.6)	-	-	-	-	25.7	20.8	70.1	614.9
20	30	100	0.5 (0.3-0.7) (0.3-0.7)	0.5 (0.1-0.7) (0.3-0.9)	4.4 (3.9-5.0) (3.9-5.2)	9.7 (8.8-10.2) (8.0-11.0)	-	-	-	-	24.6 (24.2-25.0)	17.2 (16.6-17.9)	86.4	399.9
30	30	100	0.1 (0-0.2)	0.1 (0-0.2)	3.8 (3.5-4.1)	6.8 (6.5-7.0)	-	-	-	-	22.3	13.3	98.6	366.5
20	20	100	0.6 (0.5-0.9)	0.6 (0.4-0.9)	5.2 (4.6-5.8)	11.0 (10.5-11.5)	-	-	-	-	24.3	17.7	85.7	484.7
20	35	100	0.5 (0.4-0.7)	0.5 (0.3-0.6)	4.1 (3.6-4.6)	9.1 (8.7-9.5)	-	-	-	-	24.7	17.0	86.2	369.6
20	40	100	0.5 (0.3-0.6)	0.5 (0.3-0.6)	3.8 (3.4-4.2)	8.5 (8.1-8.9)	-	-	-	-	24.9	16.8	85.6	344.8
20	30	0	0.6 (0.5-0.9)	1.3 (1.1-1.5)	4.9 (4.3-5.3)	9.9 (9.4-10.5)	-	-	-	-	27.7	19.8	79.5	463.3
20	30	33	0.6 (0.5-0.8)	0.9 (0.7-1.2)	4.7 (4.2-5.2)	10.2 (9.6-10.5)	-	-	-	-	26.7	19.0	82.4	440.9
20	30	66	0.6 (0.5-0.8)	0.6 (0.4-0.8)	4.5 (4.0-5.0)	10.1 (9.6-10.5)	-	-	-	-	25.7	18.2	85.3	419.3

Mig = 100×(prevalence of culture-positive tuberculosis in new recruits)/(prevalence of culture positive tuberculosis in the final prevalence survey)

* The values in the column labelled “Exogenous” refer to the rates of disease onset in the first year after reinfection

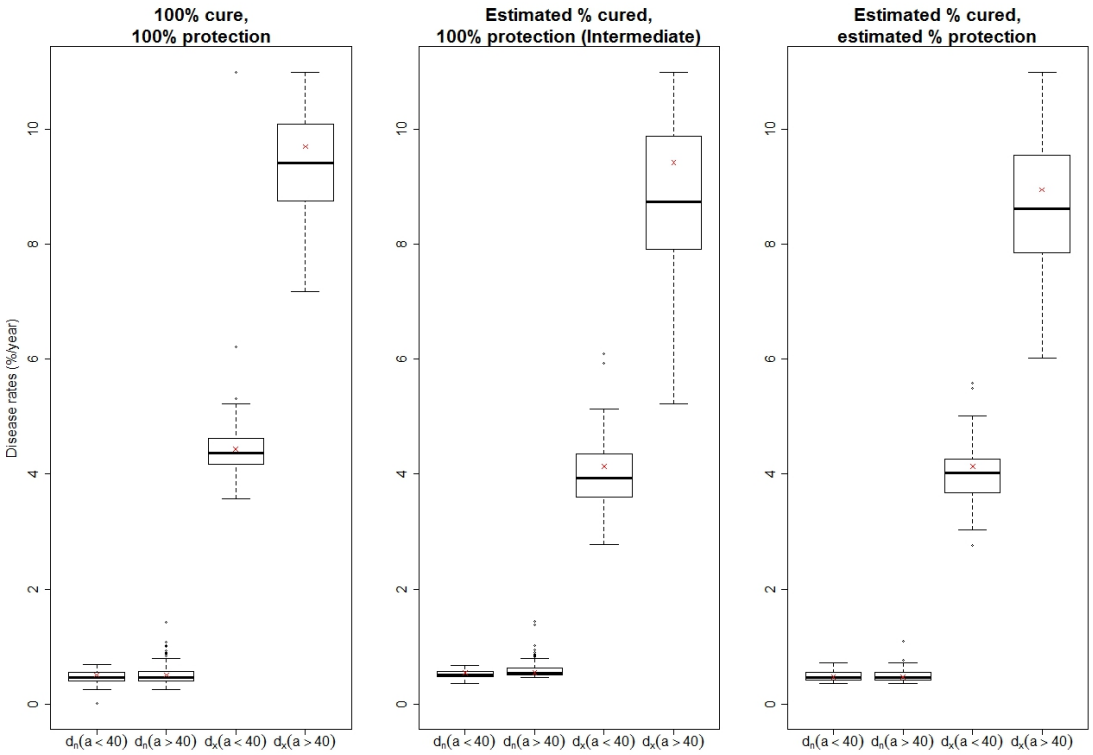
ARI (%)	HIV (%)	Mig (%)	Disease rates (HIV negative) (%/year)				% cured		%protected		Impacts		% due to reinfn	Deviance
			Reactivation (d_n)		Exogenous* (d_x)									
			Age<40	Age≥40	Age<40	Age≥40	HIV-	HIV+	HIV-	HIV+	Incidence	Prevalence		
IPT assumptions #2 (Estimated % cured, 100% protection)														
10	30	100	1.1 (0.8-1.4)	1.4 (1.4-2.0)	10.8 (8.9-11.6)	11.0 (10.8-11.6)	0-100	$4.4x10^{-3}$ (0-30.9)	-	-	18.6	10.3	70.5	579.9
20	30	100	0.5 (0.3-0.7) (0.4-0.6)	0.5 (0.4-0.7) (0.4-0.9)	4.1 (3.6-4.6) (3.1-4.9)	9.4 (7.9-9.9) (5.8-11.0)	0-100	$1.5x10^{-3}$ (0-28.0) (0-0.2)	-	-	18.1 (15.0-21.3)	10.7 (7.8-16.0)	84.7	382.3
30	30	100	0.2 (0.1-0.4)	0.2 (0.1-0.4)	3.4 (3.1-3.7)	6.4 (6.2-6.7)	0-100	$6.2x10^{-2}$ (0-35.1)	-	-	17.6	9.8	93.1	353.6
20	20	100	0.6 (0.5-0.9)	0.6 (0.4-0.9)	4.9 (4.3-5.5)	10.8 (10.3-11.3)	0-100	$1.0x10^{-3}$ (0-38.1)	-	-	18.5	11.4	84.6	469.8
20	35	100	0.5 (0.4-0.7)	0.5 (0.3-0.7)	3.8 (3.3-4.3)	8.8 (8.4-9.2)	0-100	$3.4x10^{-3}$ (0-25.4)	-	-	17.9	10.4	84.5	350.4
20	40	100	0.5 (0.4-0.6)	0.5 (0.3-0.6)	3.6 (3.0-4.0)	8.2 (7.8-8.6)	0-100	$1.7x10^{-3}$ (0-23.3)	-	-	17.6	9.9	84.2	324.3
20	30	0	0.6 (0.5-0.8)	1.2 (1.1-1.5)	4.7 (4.2-5.2)	9.6 (9.0-10.1)	0-100	$4.7x10^{-3}$ (0-22.1)	-	-	19.3	10.5	78.4	439.3
20	30	33	0.6 (0.5-0.8)	1.2 (0.9-1.4)	4.5 (3.9-5.0)	9.1 (8.6-9.6)	0-100	$4.9x10^{-3}$ (0-21.9)	-	-	18.3	9.9	77.8	419.0
20	30	66	0.6 (0.5-0.8)	0.8 (0.6-1.0)	4.4 (3.8-4.8)	9.5 (8.9-9.9)	0-100	$9.6x10^{-3}$ (0-25.6)	-	-	18.5	10.7	82.3	400.3

Mig = 100×(prevalence of culture-positive tuberculosis in new recruits)/(prevalence of culture positive tuberculosis in the final prevalence survey)
* The values in the column labelled “Exogenous” refer to the rates of disease onset in the first year after reinfection

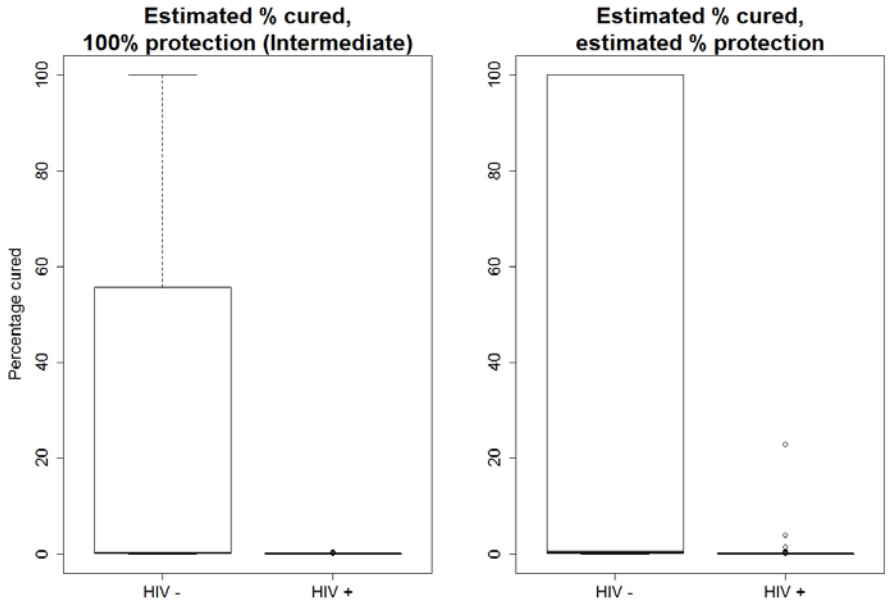
ARI (%)	HIV (%)	Mig (%)	Disease rates (HIV negative) (%/year)				% cured		%protected		Impacts		% due to reinfn	Deviance
			Reactivation (d_n)		Exogenous (d_x)									
			Age<40	Age≥40	Age<40	Age≥40	HIV-	HIV+	HIV-	HIV+	Incidence	Prevalence		
IPT assumptions #3 (Estimated % cured, estimated % protection)														
10	30	100	0.9 (0.6-1.2)	1.4 (1.3-1.9)	11.0 (9.4-11.6)	11.0 (10.8-11.6)	0-100	7.4×10^{-2} (0-25.2)	0-100	1.4×10^{-3} (0-15.7)	11.9	7.4	71.2	545.0
20	30	100	0.5 (0.3-0.7) (0.4-0.6)	0.5 (0.2-0.6) (0.4-0.6)	4.1 (3.3-4.7) (3.1-4.8)	8.9 (8.1-9.5) (6.8-10.8)	0-100	1.1×10^{-2} (0-32.6) (0-0.5)	0-100	1.2×10^{-1} (0-28.0) (0-0.6)	11.3 (10.7-15.6)	7.3 (6.7-13.9)	85.2	358.0
30	30	100	0.1 (0.0-0.3)	0.1 (0.0-0.2)	3.6 (3.3-3.9)	6.5 (6.2-6.7)	0-100	3.7×10^{-2} (0-53.5)	0-100	4.3×10^{-3} (0-29.3)	11.6	7.3	96.6	334.7
20	20	100	0.6 (0.4-0.8)	0.6 (0.3-0.7)	5.1 (4.5-5.7)	10.8 (10.3-11.3)	0-100	7.5×10^{-2} (0-39.6)	0-100	4.5×10^{-3} (0-28.5)	12.2	8.2	85.9	448.3
20	35	100	0.5 (0.3-0.6)	0.5 (0.3-0.6)	3.6 (3.2-4.1)	8.0 (7.7-8.4)	0-100	2.2×10^{-2} (0-29.1)	0-100	3.5×10^{-2} (0-27.6)	11.7	8.4	85.0	324.4
20	40	100	0.4 (0.3-0.6)	0.4 (0.3-0.6)	3.4 (2.9-3.8)	7.5 (7.2-7.8)	0-100	2.1×10^{-3} (0-26.5)	0-100	2.8×10^{-3} (0-27.0)	12.1	9.5	84.3	299.9
20	30	0	0.6 (0.5-0.8)	1.2 (1.0-1.5)	4.7 (4.2-5.2)	9.5 (8.9-10.0)	0-100	6.4×10^{-3} (0-23.7)	0-100	3.1×10^{-3} (0-23.3)	14.4	11.1	78.2	416.9
20	30	33	0.6 (0.5-0.8))	1.0 (0.8-1.2)	4.5 (4.0-5.0)	9.4 (8.9-9.9)	0-100	7.9×10^{-3} (0-25.5)	0-100	5.2×10^{-2} (0-24.3)	13.6	10.4	79.9	397.6
20	30	66	0.5 (0.4-0.7)	0.5 (0.4-0.7)	4.1 (3.6-4.5)	9.0 (8.6-9.4)	0-100	1.6×10^{-2} (0-30.0)	0-100	1.2×10^{-2} (0-26.7)	12.1	8.3	84.9	369.0

Mig = 100×(prevalence of culture-positive tuberculosis in new recruits)/(prevalence of culture positive tuberculosis in the final prevalence survey)

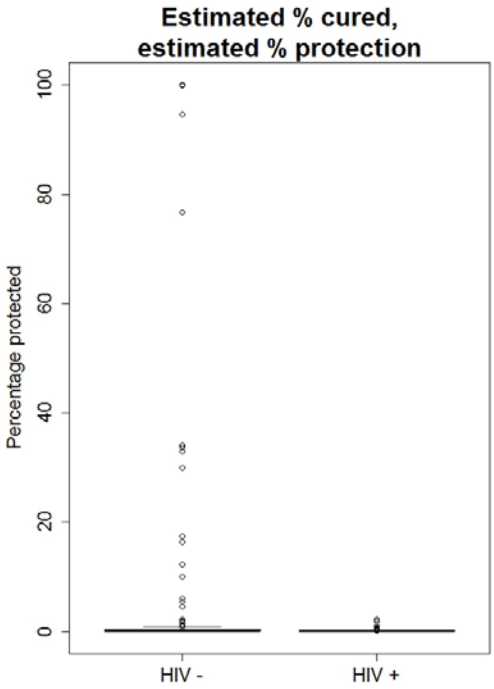
* The values in the column labelled “Exogenous” refer to the rates of disease onset in the first year after reinfection



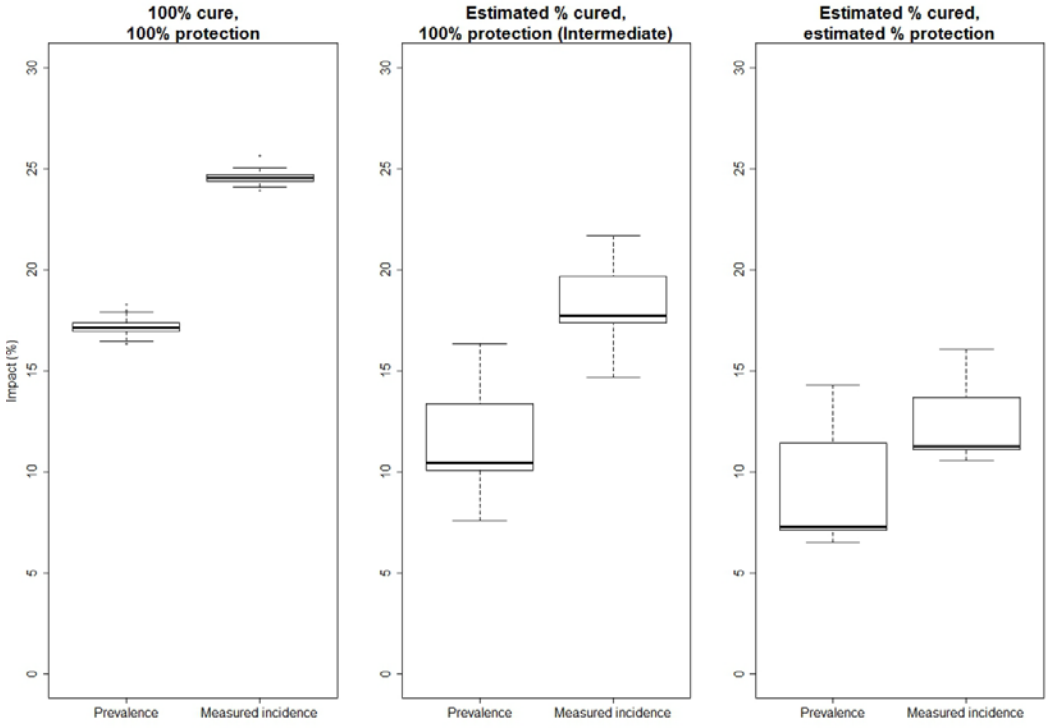
Web Figure 10: Box plots of the best-fitting values for the rates of disease onset among HIV-negative miners without radiologically-confirmed silicosis, obtained by fitting the models with the three IPT assumptions to the bootstrap data. $d_n(a<40)$ and $d_n(a\geq40)$ reflect the annual reactivation rates for those aged <40 and ≥40 years respectively. $d_x(a<40)$ and $d_x(a\geq40)$ reflect the rate of disease onset in the first year following reinfection for those aged <40 and ≥40 years respectively. The boxes reflect the interquartile range (IR), the "whiskers" extend to 1.5 times the IR and the points outside this range are represented with circles. The cross indicates the best-fitting values obtained by fitting the models to the point estimates for the incidence rate ratio and prevalence ratio.



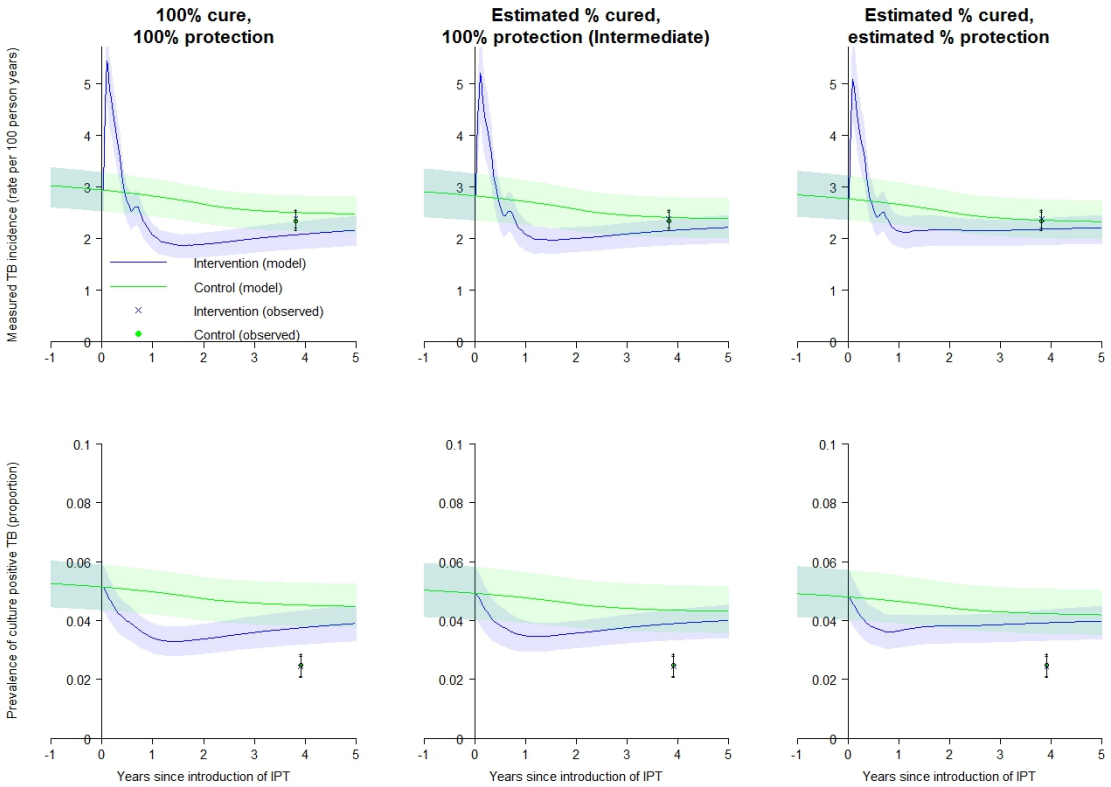
Web Figure 11: Box plot of the best-fitting values for the proportion of infections that are cured by 6 months of IPT among HIV-negative and HIV-positives, obtained by fitting the model using IPT assumptions #2 and #3 (Estimated % cured, 100% protection and Estimated % cured, estimated % protection). Web Figure 10 provides further details about the ranges indicated by the boxes, whiskers and circles.



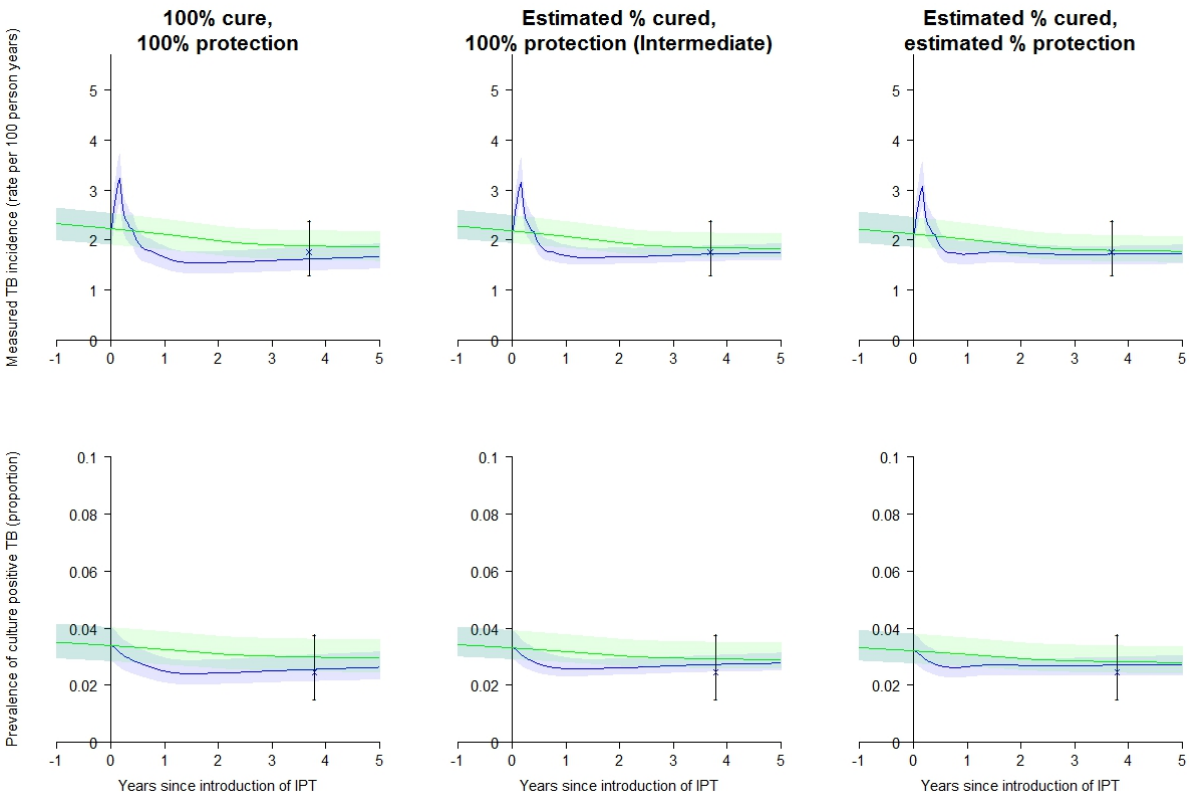
Web Figure 12: Box plot of the best-fitting values for the protection provided against reinfection whilst miners are on IPT, obtained by fitting the model using IPT assumptions #3 (Estimated % cured, estimated % protection) to the bootstrap data. See Web Figure 10 for further details about the ranges indicated by the boxes, whiskers and circles.



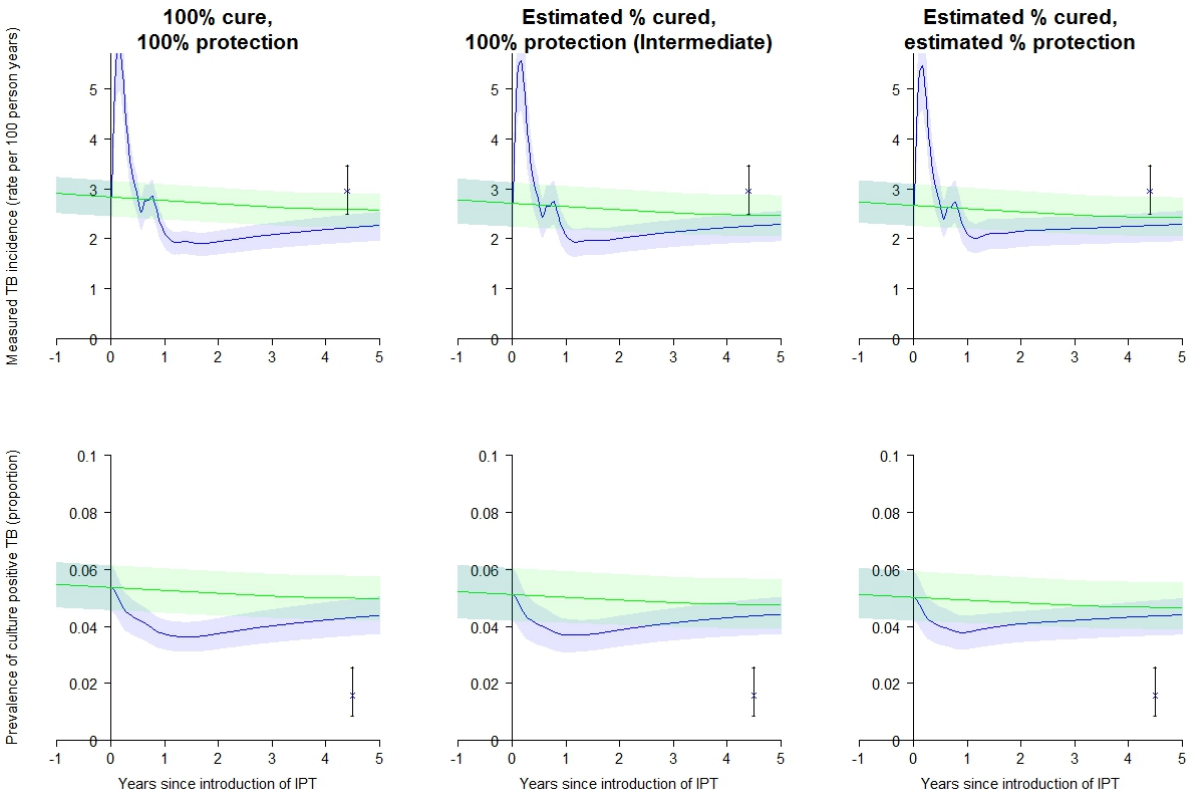
Web Figure 13: Box plot of the best-fitting impact on the measured incidence and prevalence during the primary measurement period and final prevalence survey resulting from fitting the model for all three IPT assumptions to the bootstrap data. See Web Figure 10 for further details about the ranges indicated by the boxes, whiskers and circles.



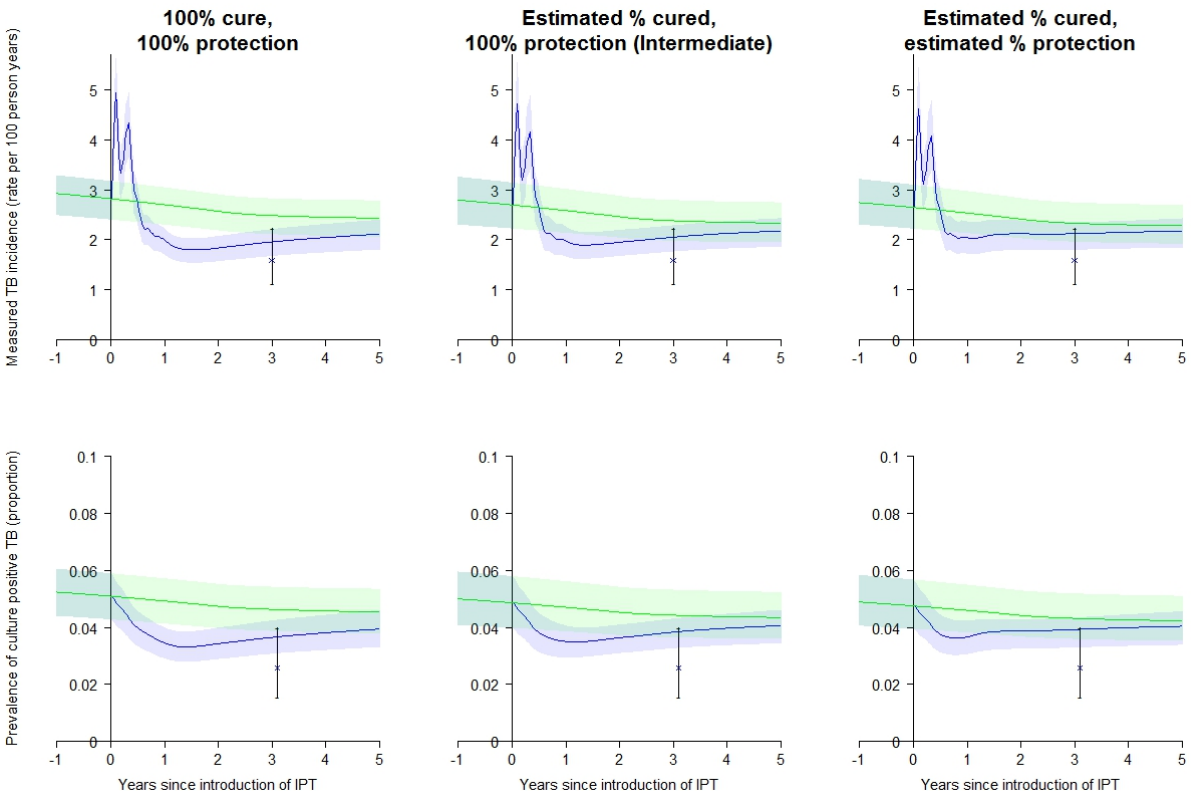
Web Figure 14a: Summary of the best-fitting values for the TB incidence and prevalence spanned by the best-fitting models to the bootstrap datasets, considering all intervention clusters combined. The shaded areas reflect the 95% range of the best-fitting incidence (top row) and prevalence (bottom row). Figures b-i show the corresponding cluster-specific plots.



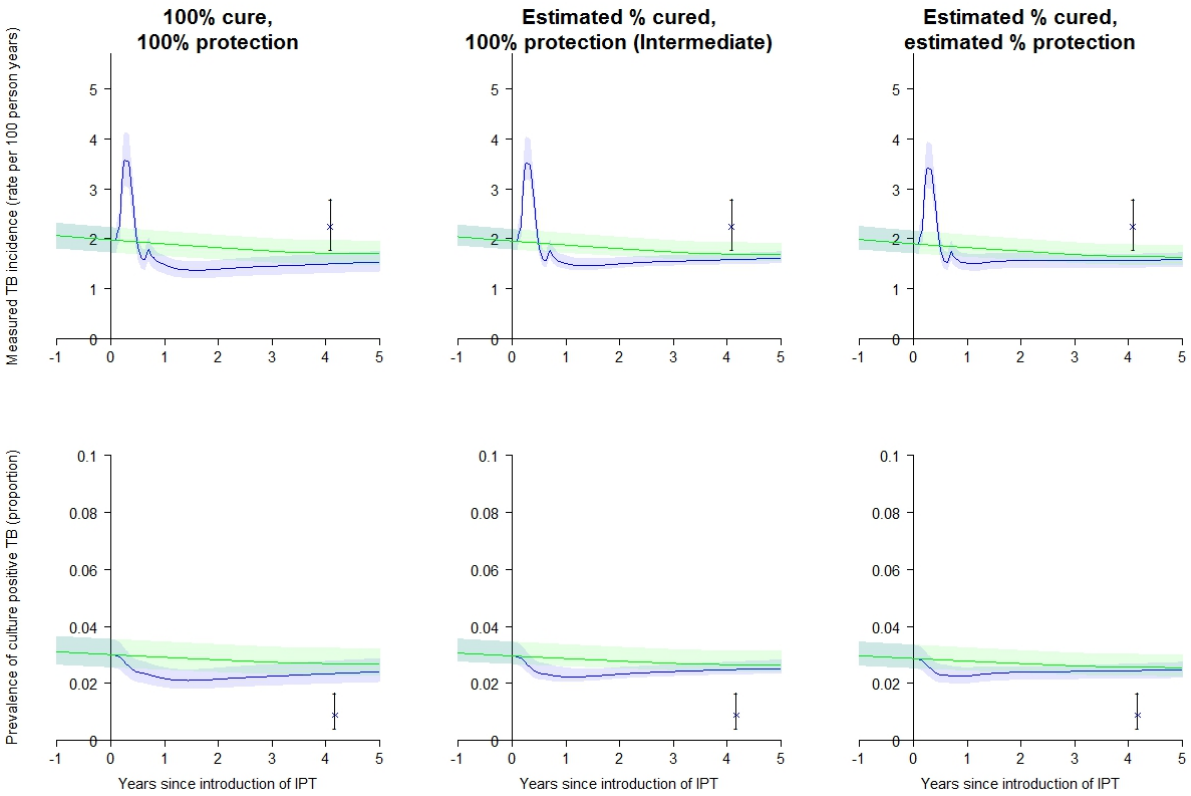
Web Figure 14b: As for Web Figure 14a but considering cluster 1.



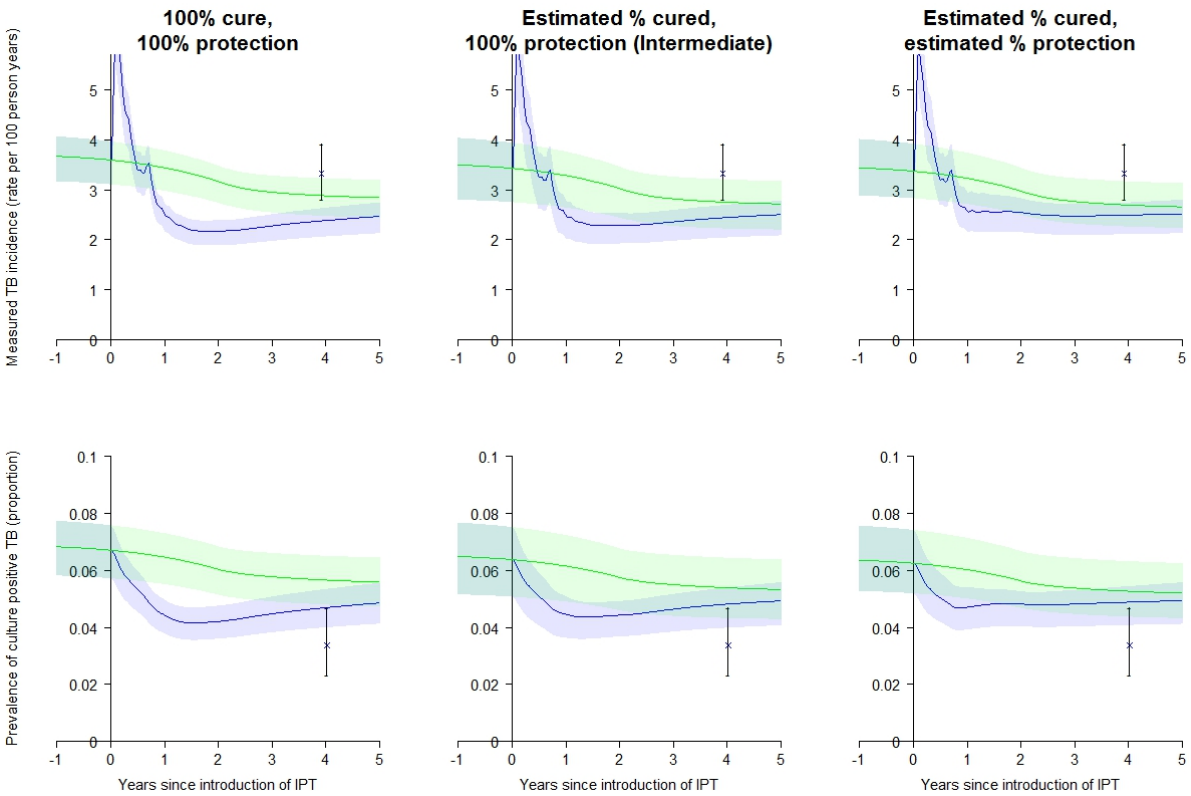
Web Figure 14c: As for Web Figure 14a but considering cluster 2.



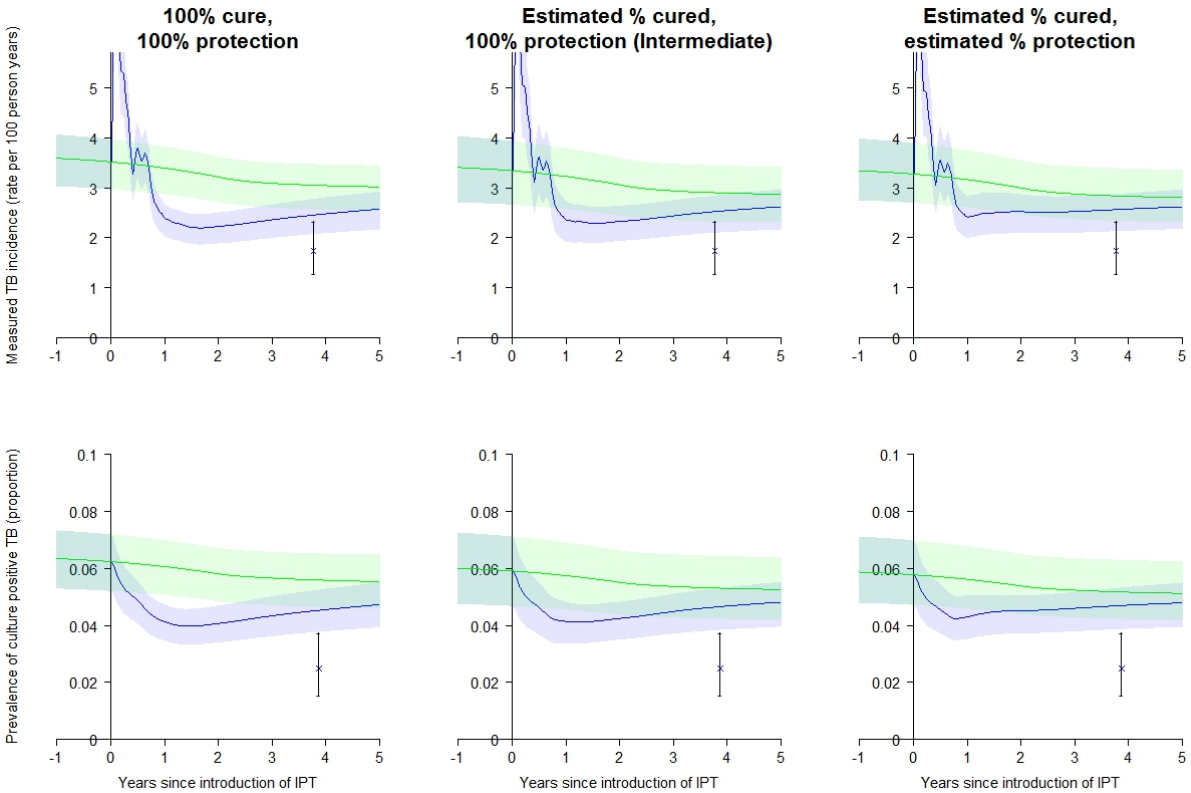
Web Figure 14d: As for Web Figure 14a but considering cluster 3.



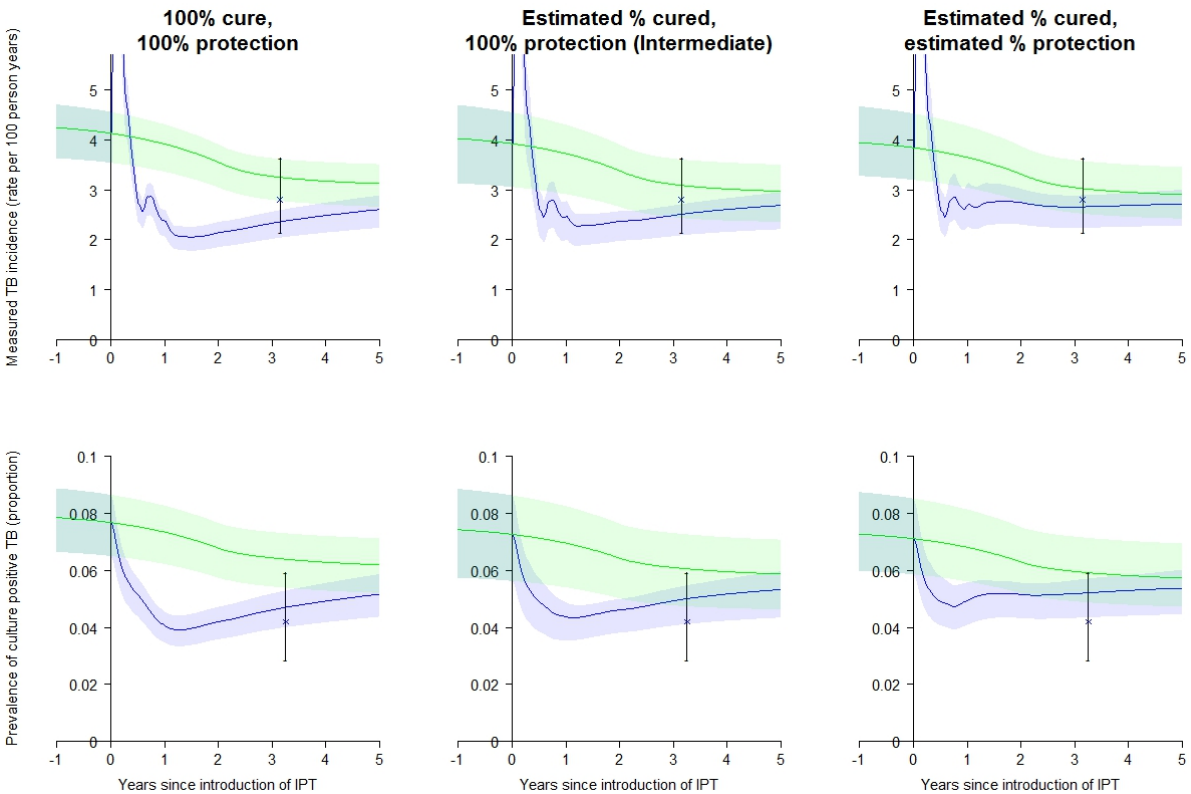
Web Figure 14e: As for Web Figure 14a, but considering cluster 4.



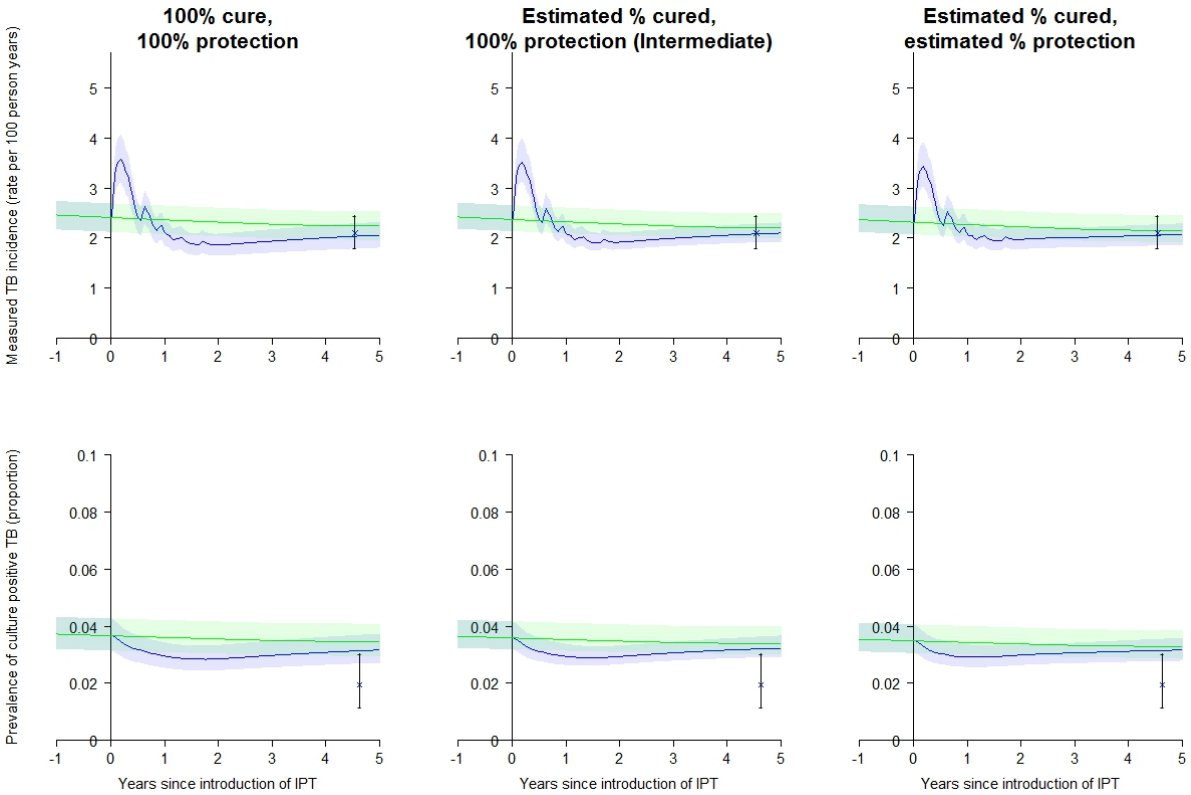
Web Figure 14f: As for Web Figure 14a, but considering cluster 5.



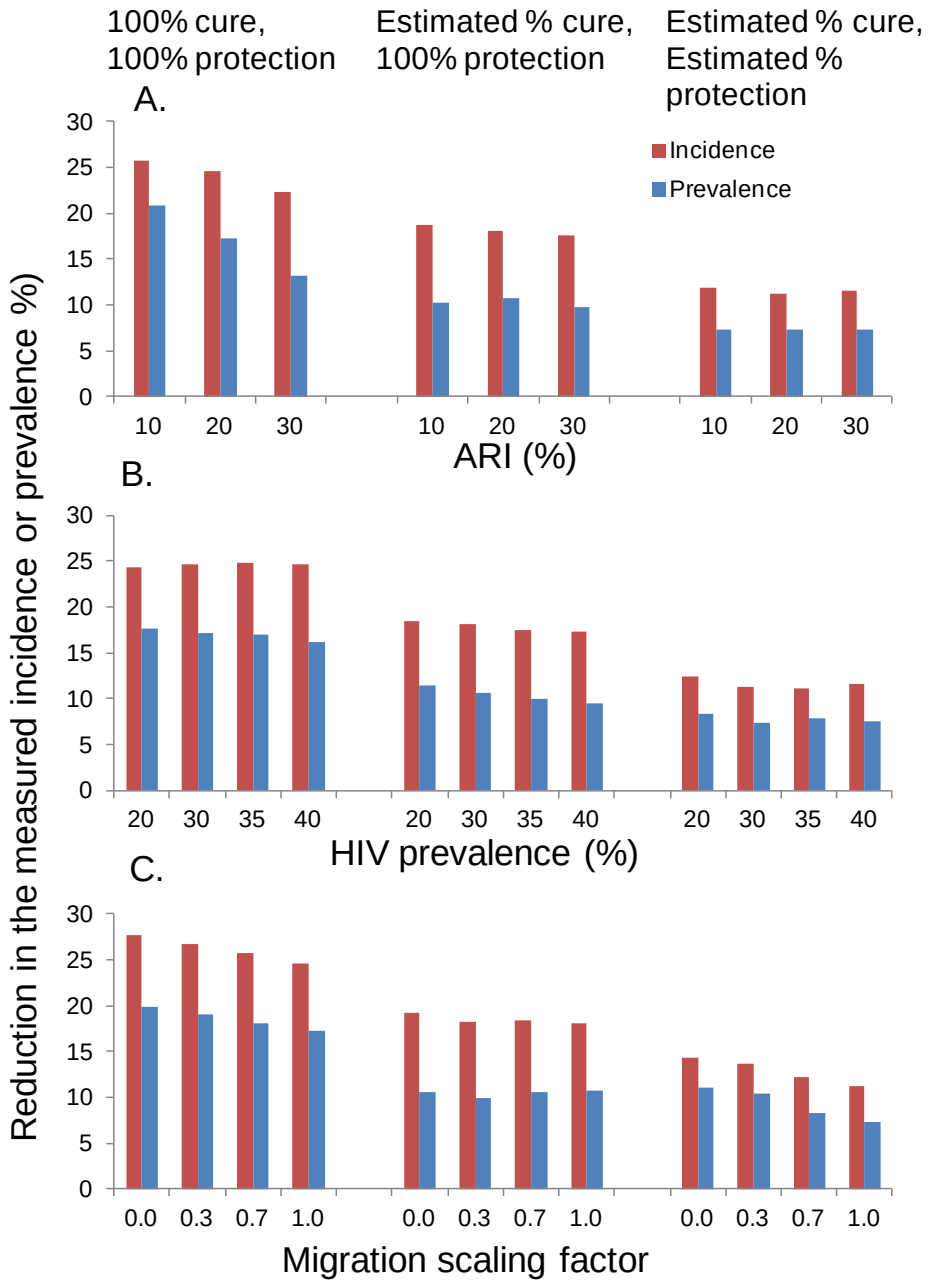
Web Figure 14g: As for Web Figure 14a, but considering cluster 6.



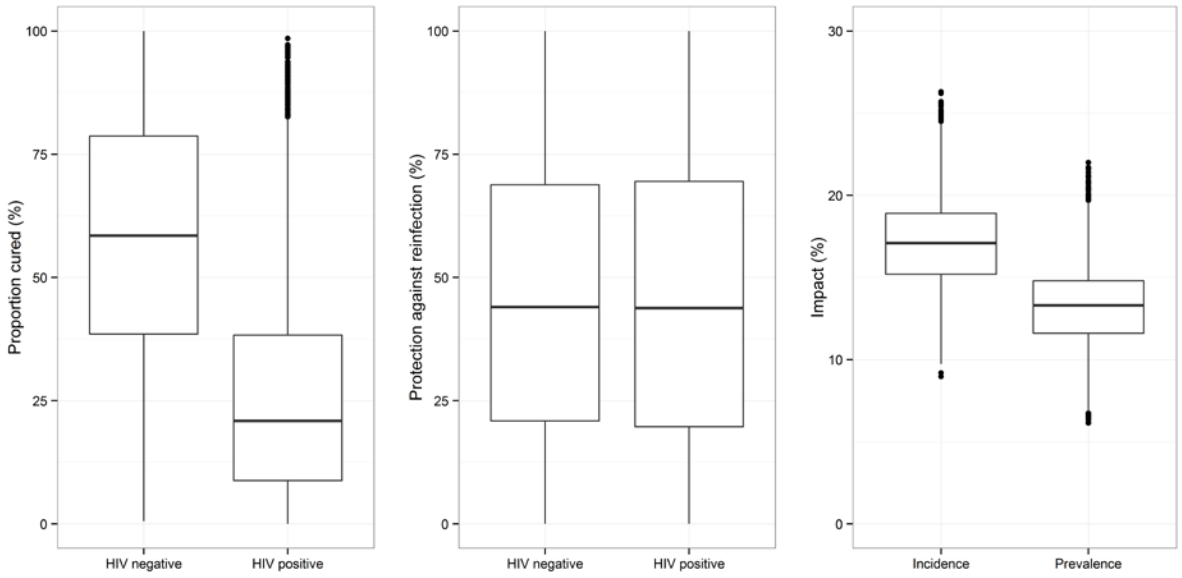
Web Figure 14h: As for Web Figure 14a, but considering cluster 7.



Web Figure 14i: As for Web Figure 14a, but considering cluster 8.

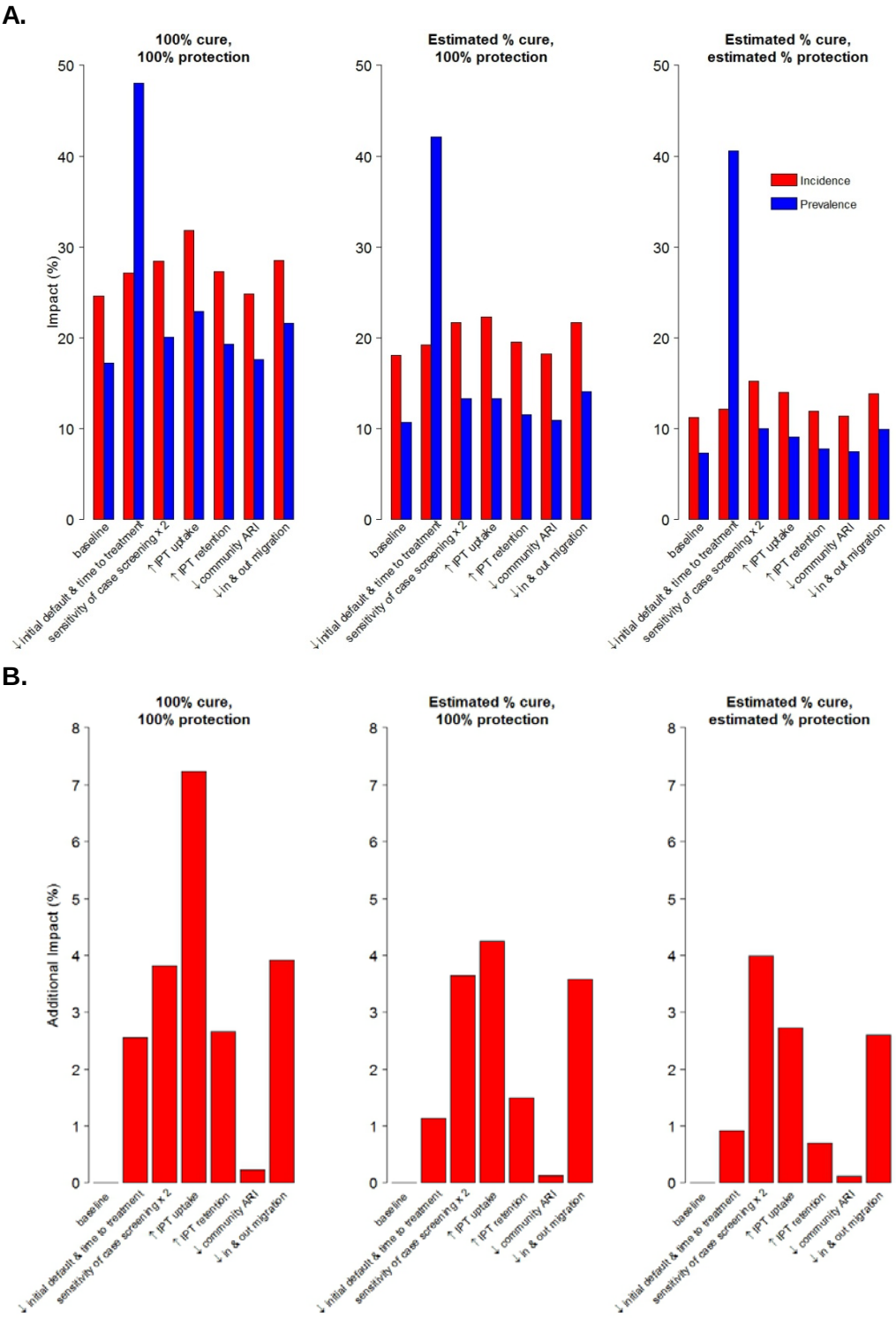


Web Figure 15: Sensitivity of the best-fitting impact on the intervention obtained for the three IPT assumptions on the measured incidence and prevalence during the primary measurement periods and final prevalence survey to the assumed A. annual risk of *M tuberculosis* infection, B. HIV prevalence and C. the factor by which the TB prevalence among new recruits differs from that estimated in the final prevalence survey ("migration scaling factor").

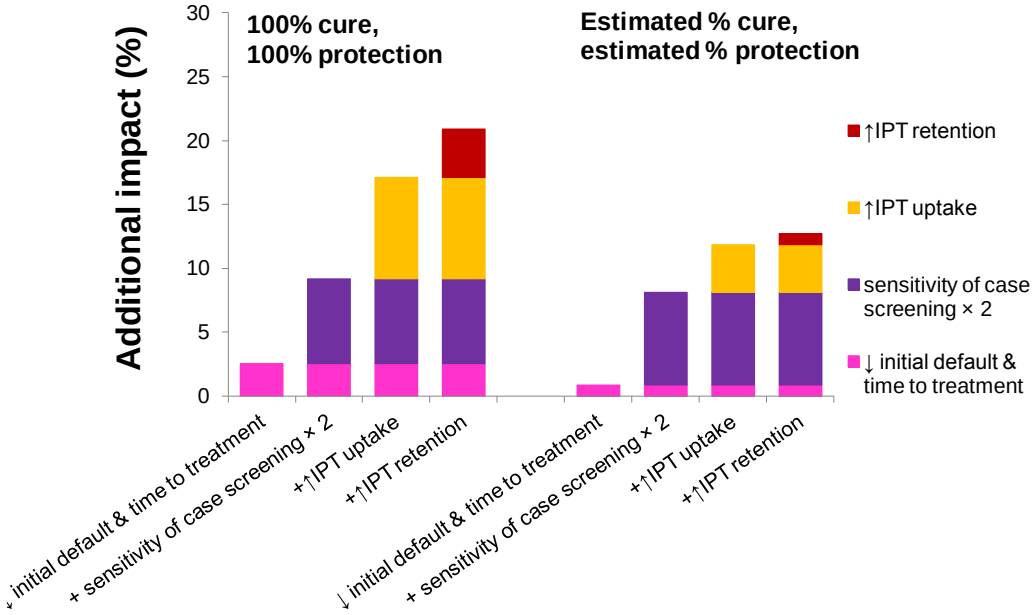


Web Figure 16: Results of the Bayesian melding (resampling 20,000 parameter combinations from 2.28 million parameter combinations using the likelihood of the prevalence as the weight). Box plot of estimates of the proportion of infections that were cured by 6 months of IPT, the protection provided by IPT against reinfection and the impact of the intervention. The boxes reflect the interquartile range (IR), the "whiskers" extend to 1.5 times the IR and the points outside this range are represented with circles. The resampling process resulted in 8728 unique parameter combinations.

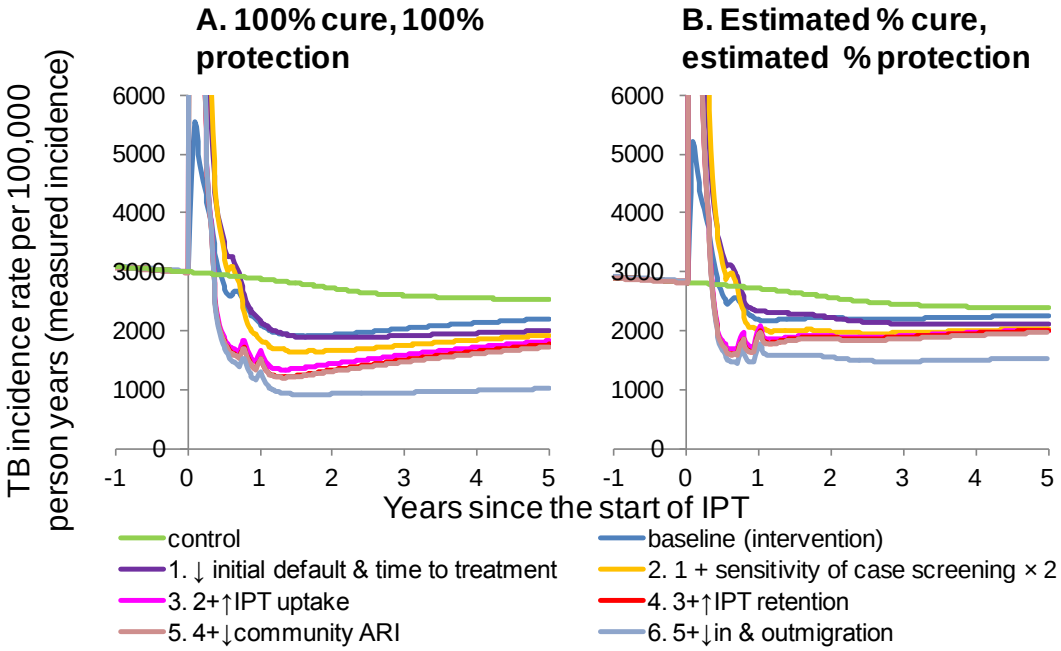
5.2 Objective 2: What was the maximum impact Thibela TB could have achieved with optimal implementation?



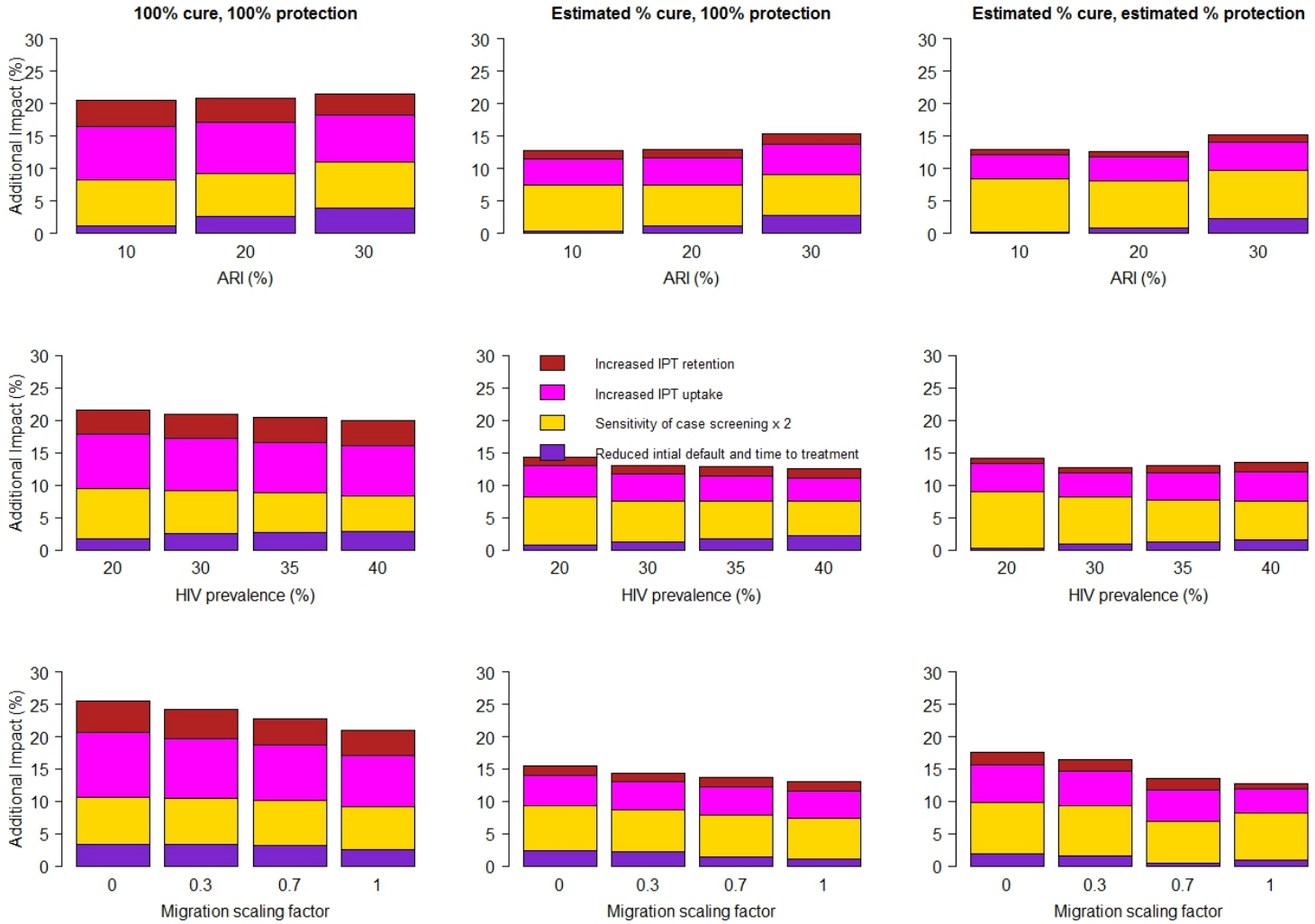
Web Figure 17: Predicted effect of changing factors individually on the impact in the measured incidence for the best-fitting base-case models, for the three IPT assumptions. The assumed prevalence of culture-positive TB among new employees equals that seen in the intervention clusters in the final prevalence survey (after dividing by 0.98).



Web Figure 18: Predicted effect of changing individual factors incrementally on the impact on the measured incidence in the best-fitting models, for IPT assumptions #1 and #3 (100% cured, 100% protection and Estimated % cured, estimated % protection respectively). The assumed prevalence of culture-positive TB among new employees equals that seen in the intervention clusters in the final prevalence survey (after dividing by 0.98).

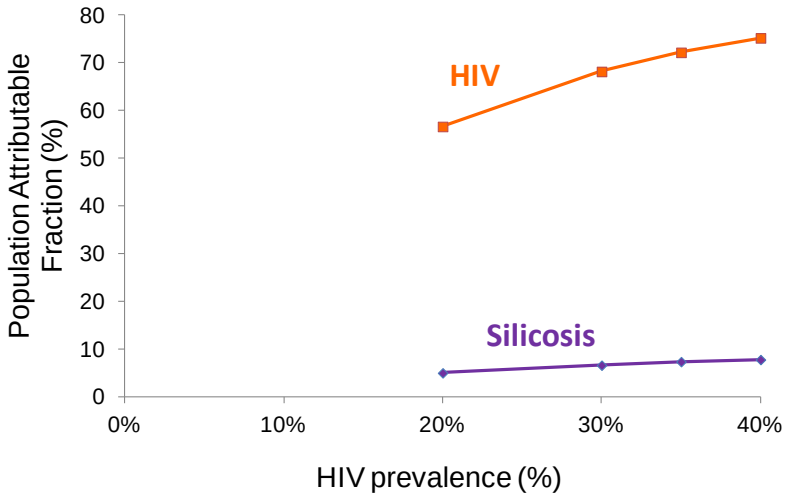


Web Figure 19: Predicted effect of optimising the intervention. Predicted effect of changing individual factors incrementally on the impact on the measured incidence in the best-fitting models, for IPT assumptions #1 and #3 (100% cured, 100% protection and Estimated % cured, estimated % protection respectively). Option 5 includes the effect of removing contact with the community outside the mines (↓ community ARI); option 6 includes the additional effect of reducing in and outmigration by 75%. The assumed prevalence of culture-positive TB among new employees equals that seen in the intervention clusters in the final prevalence survey (after dividing by 0.98).

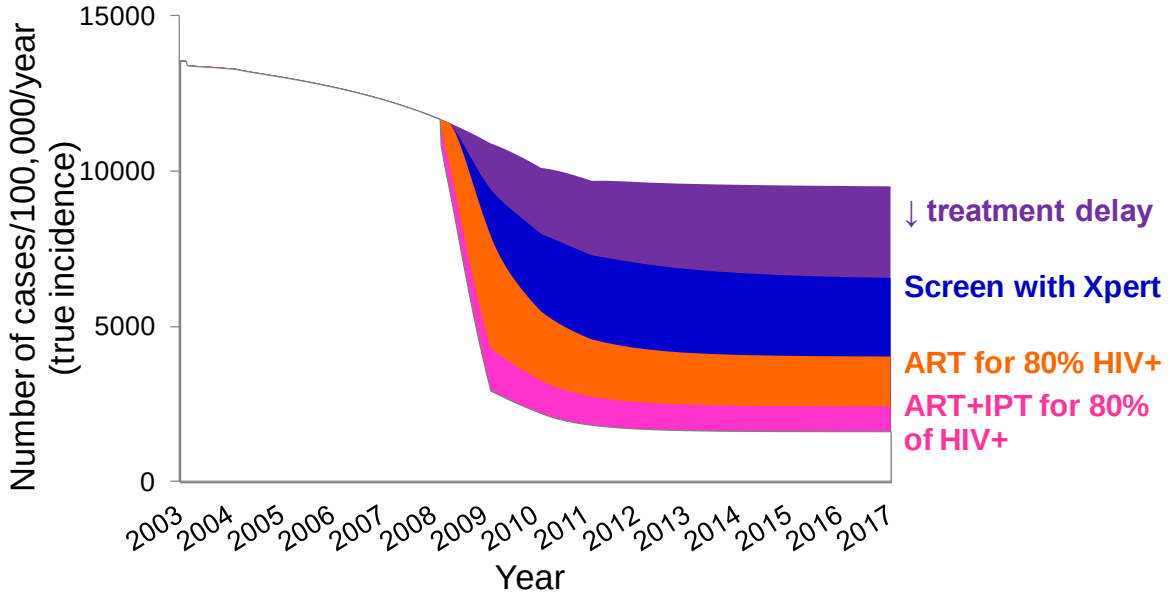


Web Figure 20: Sensitivity of the predicted effect of optimising the intervention to the ARI, HIV prevalence and the factor by which the TB prevalence among newly employed miners differs from that inferred in the final prevalence survey (migration scaling factor=0 implies that the prevalence was zero among newly employed miners). The assumed rates of disease onset, proportion of latent infections cured and protection against reinfection are those obtained after fitting the corresponding model (Web Table 12).

5.3 Objective 3: What might control tuberculosis in goldmines?



Web Figure 21: The population attributable fraction (PAF) of the TB incidence from HIV and silicosis for all clusters. As described in the main text, the PAF for silicosis greatly underestimates the reduction in tuberculosis incidence which might result from dust control. This plot was similar for all clusters and generally insensitive to the assumed ARI and the prevalence of culture-positive TB among new mining employees



Web Figure 22: Predicted effect of combining interventions on the true tuberculosis incidence among HIV-positives. Predicted impact of introducing reduced treatment delay, screening with Xpert MTB/RIF, ART for 80% of HIV-positives and IPT to those on ART. The shaded areas show the incremental effect of adding the intervention considered. For the scenario involving Xpert MTB/RIF, Xpert MTB/RIF is used in routine medical examination, for newly employed miners and on passive presentation. For both the ART and ART/IPT scenarios, the coverage is increased to reach 80% by 2009.

References

1. R. Wood, H. Liang, H. Wu, K. Middelkoop, T. Oni, M. X. Rangaka, R. J. Wilkinson, L. G. Bekker, S. D. Lawn, Changing prevalence of tuberculosis infection with increasing age in high-burden townships in South Africa. *Int J Tuberc Lung Dis* **14**, 406-412 (2010).
2. M. A. Behr, S. A. Warren, H. Salamon, P. C. Hopewell, A. Ponce de Leon, C. L. Daley, P. M. Small, Transmission of Mycobacterium tuberculosis from patients smear-negative for acid-fast bacilli. *Lancet* **353**, 444-449 (1999).
3. H. Abbey, An examination of the Reed-Frost theory of epidemics. *Hum Biol* **24**, 201-233 (1952).
4. J. G. Laing, Tuberculosis in the mining industry. *Proc Mine Med Off Assoc* **48**, 8-19 (1968).
5. E. L. Corbett, K. F. Mallory, A. D. Grant, G. J. Churchyard, K. M. De Cock, HIV-1 infection and risk of tuberculosis after rifampicin treatment. *Lancet* **357**, 957-958 (2001).
6. P. Sonnenberg, J. Murray, J. R. Glynn, S. Shearer, B. Kambashi, P. Godfrey-Faussett, HIV-1 and recurrence, relapse, and reinfection of tuberculosis after cure: a cohort study in South African mineworkers. *Lancet* **358**, 1687-1693 (2001).
7. E. L. Corbett, G. J. Churchyard, T. C. Clayton, B. G. Williams, D. Mulder, R. J. Hayes, K. M. De Cock, HIV infection and silicosis: the impact of two potent risk factors on the incidence of mycobacterial disease in South African miners. *AIDS* **14**, 2759-2768 (2000).
8. K. Styblo, *Epidemiology of tuberculosis. Selected Papers*. (KNCV, The Hague, 1991).
9. I. Sutherland, "The ten-year incidence of clinical tuberculosis following "conversion" in 2,550 individuals aged 14 to 19 years. ," *TSRU Progress Report* (KNCV, The Hague, The Netherlands, 1968).
10. I. Sutherland, E. Svandova, S. Radhakrishna, The development of clinical tuberculosis following infection with tubercle bacilli. 1. A theoretical model for the development of clinical tuberculosis following infection, linking from data on the risk of tuberculous infection and the incidence of clinical tuberculosis in the Netherlands. *Tubercle* **63**, 255-268 (1982).
11. E. Vynnycky, P. E. Fine, The natural history of tuberculosis: the implications of age-dependent risks of disease and the role of reinfection. *Epidemiol Infect* **119**, 183-201 (1997).
12. M. Clark, E. Vynnycky, The use of maximum likelihood methods to estimate the risk of tuberculous infection and disease in a Canadian First Nations population. *Int J Epidemiol* **33**, 477-484 (2004).
13. E. Vynnycky, M. W. Borgdorff, C. C. Leung, C. M. Tam, P. E. Fine, Limited impact of tuberculosis control in Hong Kong: attributable to high risks of reactivation disease. *Epidemiol Infect* **136**, 943-952 (2008).
14. B. G. Williams, R. Granich, K. M. De Cock, P. Glaziou, A. Sharma, C. Dye, Antiretroviral therapy for tuberculosis control in nine African countries. *Proc Natl Acad Sci U S A* **107**, 19485-19489 (2010).
15. E. L. Corbett, S. Charalambous, V. M. Moloi, K. Fielding, A. D. Grant, C. Dye, K. M. De Cock, R. J. Hayes, B. G. Williams, G. J. Churchyard, Human immunodeficiency virus and the prevalence of undiagnosed tuberculosis in African gold miners. *Am J Respir Crit Care Med* **170**, 673-679 (2004).
16. J. J. Lewis, S. Charalambous, J. H. Day, K. L. Fielding, A. D. Grant, R. J. Hayes, E. L. Corbett, G. J. Churchyard, HIV infection does not affect active case finding of tuberculosis in South African gold miners. *Am J Respir Crit Care Med* **180**, 1271-1278 (2009).

17. B. G. Williams, E. L. Korenromp, E. Gouws, G. P. Schmid, B. Auvert, C. Dye, HIV infection, antiretroviral therapy, and CD4+ cell count distributions in African populations. *J Infect Dis* **194**, 1450-1458 (2006).
18. A. B. Suthar, S. D. Lawn, J. del Amo, H. Getahun, C. Dye, D. Sculier, T. R. Sterling, R. E. Chaisson, B. G. Williams, A. D. Harries, R. M. Granich, Antiretroviral therapy for prevention of tuberculosis in adults with HIV: a systematic review and meta-analysis. *PLoS Med* **9**, e1001270 (2012).
19. G. J. Churchyard, K. Fielding, S. Roux, E. L. Corbett, R. E. Chaisson, K. M. De Cock, R. J. Hayes, A. D. Grant, Twelve-monthly versus six-monthly radiological screening for active case-finding of tuberculosis: a randomised controlled trial. *Thorax* **66**, 134-139 (2011).
20. G. J. Churchyard, K. L. Fielding, J. J. Lewis, L. Coetzee, E. L. Corbett, P. Godfrey-Faussett, R. J. Hayes, R. E. Chaisson, A. D. Grant, A trial of mass isoniazid preventive therapy for tuberculosis control. *New England Journal of Medicine* **370**, 301-310 (2013).
21. T. Samandari, T. B. Agizew, S. Nyirenda, Z. Tedla, T. Sibanda, N. Shang, B. Mosimaneotsile, O. I. Motsamai, L. Bozeman, M. K. Davis, E. A. Talbot, T. L. Moeti, H. J. Moffat, P. H. Kilmarx, K. G. Castro, C. D. Wells, 6-month versus 36-month isoniazid preventive treatment for tuberculosis in adults with HIV infection in Botswana: a randomised, double-blind, placebo-controlled trial. *Lancet* **377**, 1588-1598 (2011).
22. G. W. Comstock, Isoniazid prophylaxis in an undeveloped area. *Am Rev Respir Dis* **86**, 810-822 (1962).
23. L. E. Dodd, R. J. Wilkinson, Isoniazid preventive therapy in HIV infection. *Lancet* **377**, 1548-1550 (2011).
24. M. X. Rangaka, A. Boulle, R. J. Wilkinson, G. van Cutsem, E. Goemaere, R. Goliath, R. Titus, S. Mathee, G. Maartens, in *AIDS 2012. XIX International AIDS Conference*. (Washington DC, 2012).
25. J. E. Golub, P. Pronyk, L. Mohapi, N. Thsabangu, M. Moshabela, H. Struthers, G. E. Gray, J. A. McIntyre, R. E. Chaisson, N. A. Martinson, Isoniazid preventive therapy, HAART and tuberculosis risk in HIV-infected adults in South Africa: a prospective cohort. *AIDS* **23**, 631-636 (2009).
26. J. E. Golub, V. Saraceni, S. C. Cavalcante, A. G. Pacheco, L. H. Moulton, B. S. King, A. Efron, R. D. Moore, R. E. Chaisson, B. Durovni, The impact of antiretroviral therapy and isoniazid preventive therapy on tuberculosis incidence in HIV-infected patients in Rio de Janeiro, Brazil. *AIDS* **21**, 1441-1448 (2007).
27. K. L. Fielding, A. D. Grant, R. J. Hayes, R. E. Chaisson, E. L. Corbett, G. J. Churchyard, Thibela TB: design and methods of a cluster randomised trial of the effect of community-wide isoniazid preventive therapy on tuberculosis amongst gold miners in South Africa. *Contemp Clin Trials* **32**, 382-392 (2011).
28. E. W. Tiemersma, M. J. van der Werf, M. W. Borgdorff, B. G. Williams, N. J. Nagelkerke, Natural history of tuberculosis: duration and fatality of untreated pulmonary tuberculosis in HIV negative patients: a systematic review. *PLoS One* **6**, e17601 (2011).
29. G. J. Churchyard, I. Kleinschmidt, E. L. Corbett, J. Murray, J. Smit, K. M. De Cock, Factors associated with an increased case-fatality rate in HIV-infected and non-infected South African gold miners with pulmonary tuberculosis. *Int J Tuberc Lung Dis* **4**, 705-712 (2000).
30. M. Straetemans, P. Glaziou, A. L. Bierrenbach, C. Sismanidis, M. J. van der Werf, Assessing tuberculosis case fatality ratio: a meta-analysis. *PLoS One* **6**, e20755 (2011).
31. W. H. Press, *Numerical recipes in C : the art of scientific computing*. (Cambridge University Press, ed. 2nd ed., 1992).

32. S. E. Dorman, V. N. Chihota, J. J. Lewis, M. Shah, D. Clark, A. D. Grant, G. J. Churchyard, K. L. Fielding, Performance characteristics of the Cepheid Xpert MTB/RIF test in a tuberculosis prevalence survey. *PLoS One* **7**, e43307 (2012).
33. S. D. Lawn, S. V. Brooks, K. Kranzer, M. P. Nicol, A. Whitelaw, M. Vogt, L. G. Bekker, R. Wood, Screening for HIV-associated tuberculosis and rifampicin resistance before antiretroviral therapy using the Xpert MTB/RIF assay: a prospective study. *PLoS Med* **8**, e1001067 (2011).
34. G. Theron, J. Peter, R. van Zyl-Smit, H. Mishra, E. Streicher, S. Murray, R. Dawson, A. Whitelaw, M. Hoelscher, S. Sharma, M. Pai, R. Warren, K. Dheda, Evaluation of the Xpert MTB/RIF assay for the diagnosis of pulmonary tuberculosis in a high HIV prevalence setting. *Am J Respir Crit Care Med* **184**, 132-140 (2011).
35. S. D. Lawn, M. P. Nicol, Xpert(R) MTB/RIF assay: development, evaluation and implementation of a new rapid molecular diagnostic for tuberculosis and rifampicin resistance. *Future Microbiol* **6**, 1067-1082 (2011).
36. C. C. Boehme, M. P. Nicol, P. Nabeta, J. S. Michael, E. Gotuzzo, R. Tahirli, M. T. Gler, R. Blakemore, W. Worodria, C. Gray, L. Huang, T. Caceres, R. Mehdiyev, L. Raymond, A. Whitelaw, K. Sagadevan, H. Alexander, H. Albert, F. Cobelens, H. Cox, D. Alland, M. D. Perkins, Feasibility, diagnostic accuracy, and effectiveness of decentralised use of the Xpert MTB/RIF test for diagnosis of tuberculosis and multidrug resistance: a multicentre implementation study. *Lancet* **377**, 1495-1505 (2011).
37. K. K. Orroth, R. G. White, E. L. Korenromp, R. Bakker, J. Changalucha, J. D. Habbema, R. J. Hayes, Empirical observations underestimate the proportion of human immunodeficiency virus infections attributable to sexually transmitted diseases in the Mwanza and Rakai sexually transmitted disease treatment trials: Simulation results. *Sex Transm Dis* **33**, 536-544 (2006).