Nationwide and regional decline in incidence of microbiologically-confirmed pulmonary tuberculosis in South Africa: a time series analysis from 2004 to 2012

Supplemental Appendix

TABLE OF CONTENTS

Contents

1. Supplementary Methods

1.1. Sources of data

The NHLS Corporate Data Warehouse Laboratory TB Database

All specimens from presumptive TB cases screened at public health facilities and from cases diagnosed in the private sector entering treatment in the public sector are sent to the NHLS' national network of laboratories. The majority (84%) of South Africans do not have private medical insurance and use public health facilities, including 91% of Black Africans who are also disproportionately affected by HIV (31% prevalence).[1, 2] Furthermore, individuals with TB, including those with private medical insurance, make use of public health facilities for TB management, provided at no-cost by the Department of Health.[3]

A laboratory request form with demographic data accompanies each specimen and is recorded onto the laboratory information system (LIS). Once testing is completed, all available results are captured against the demographic data. The absence of unique patient identifiers required a series of record-linking processes, to match multiple specimen records to individual patients. This was done firstly by direct linking of records where the surnames, first names and dates of birth were an exact match; followed by probabilistic matching, using an algorithm to assess the probability that a set of records are linked to the same individual. The resulting probabilities were weighted and averaged to provide a final score on which matching was based, and sequential numeric identifiers were assigned to sets of records most likely belonging to the same individuals.

The Electronic TB Database

The Electronic TB Register (ETR) exists as a series of databases per province and is not consolidated into a national database, while the Electronic Drug Resistant TB Register (EDR) is an online database housed on a web server. This data may also contain duplicates caused by patients attending multiple treatment facilities or health care workers opening up new files (and therefore new ETR/EDR records) for existing patients. For the purposes of this analysis, the provincial ETR datasets were integrated and cleaned to create a national dataset of newly

registered PTB cases recorded between 01 January 2004 and 31 December 2012. Data on new cases recorded in the EDR for the same period were extracted and combined with the ETR data to create a national Electronic TB database (ETD). We compared monthly PTB incidence rates (PTBr) from the ETD with monthly microbiologically-confirmed PTB (mPTB) incidence rates.

Strengths and limitations of the TB surveillance data used

There are challenges with the completeness of the ETD, as data for Mpumalanga province for 2004 to 2006 was lost, while there were gaps in the data made available for Gauteng and Northern Cape, and KwaZulu-Natal to a lesser extent.

Although HIV status and ART usage in TB-HIV co-infected individuals is recorded on the ETD, there was inadequate recording of this data during the early years of the implementation of the registers.

CDW provides an integrated, national dataset of all public sector diagnosed TB cases in South Africa, as well as basic demographic information. It does not however, contain clinical data and treatment history. While CDW does not capture patients diagnosed in private sector laboratories, these patients will enter treatment in public sector facilities and will appear in the ETD. Nonetheless, CDW data is a robust source of TB surveillance data as the NHLS is the sole laboratory service provider to the public sector facilities in South Africa, at which the majority of the population access care. There is no unique identifier routinely used and common to the ETD and National Health Laboratory Service Corporate Data Warehouse (CDW) data which can be used to link them on a patient-level, and errors or discrepancies in recording of names, addresses and dates of birth makes linking using these parameters a challenge.

The ASSA 2008 model

The ASSA 2008 AIDS and Demographic models were developed to provide retrospective estimates and future projections of HIV prevalence and incidence in adults and children per province in South Africa. It also provides data on ART coverage, ANC coverage, HIV-related mortality and population.

Strengths and limitations of the ASSA 2008 model

Prior to the development of the ASSA models, there is scant data on HIV and ART in the general population, as data was only recorded on this group from 2011, with periodic surveys undertaken in 2008 and 2011. As the ASSA model provides estimates, these may under-report HIV prevalence and incidence, and ART coverage at best. However, a relative strength of this data is that it is available nationally and per province from 1985 to 2025, therefore covering the period of observation for this study.

1.2. Statistical methods

To determine an appropriate method to model the data, we considered the characteristics of the data and the objective at hand. The primary objective of this paper is to describe trends in incidence of pulmonary tuberculosis nationally and provincially in South Africa, between 2004 and 2012.

To begin exploring the nature of the data, time series plots of monthly PTB incidence rates and three-month moving averages were examined. This was done at a national and provincial level and then further stratified by age group (<15, 15-24, 25-44, 45-64, ≥65). Major characteristics found were the non-stationarity of the data, parabolic shape of the underlying trend and the presence of seasonality and autocorrelation. The authors investigated two different models which accommodate autocorrelation and do not assume a stationary series: an autoregressive model (AR) up to order 12 and a nonlinear model (NL) with sinusoidal terms to capture seasonality. Linear and quadratic terms for time were included in both models to allow for the quadratic nature of the underlying trend. Models were fitted in SAS and compared using the mean square error.

A Poisson autoregressive model was initially considered. However, after visual inspection of the observed and predicted incidence rates, the model continuously showed to be overfitting the data. As our primary objective was to describe the underlying trend in mPTB, we did not consider using an S/ARIMA model after differencing the data.

AR model

The autoregressive model is as follows:

$$\begin{aligned} y_t &= \beta_0 + \beta_1 t + \beta_2 t^2 + \epsilon_t \\ \epsilon_t &= w_t - \phi_1 \epsilon_{t-1} - \ldots - \phi_{12} \epsilon_{t-12} \\ w_t &\sim N(0, \sigma^2) \end{aligned} \tag{1}$$

where *t* indicates month with a value of 1 referring to January 2004 and a value of 108 referring to December 2012, and y_t is the PTB incidence at time *t*. We refer to $\beta_0 + \beta_1 t + \beta_2 t_2$ as the structural portion of the model and ε_t as the autoregressive portion with autoregression coefficients $\varphi_1 \ldots \varphi_{12}$. The autoregressive portion of the model accommodates the seasonality and autocorrelation apparent in the data.

Depending on the province and/or age group, not all 12 autoregression coefficients were needed in the model. We used backward elimination to remove insignificant autoregressive coefficients. The autoregressive parameters included in the final model are reported in Table (i).

NL model

The nonlinear model is as follows:

$$\begin{aligned} y_t =& \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_{ampS} \sin\left(\frac{\beta_{multS} 2\pi(m_t - \beta_{lag})}{12}\right) \\ &+ \beta_{ampC} \cos\left(\frac{\beta_{multC} 2\pi(m_t - \beta_{lag})}{12}\right) + \epsilon_t \end{aligned} \tag{2}$$

where y_t and *t* are as above and m_t represents the month of the year at time *t*.

Depending on the province and/or age group, the above model would not converge.
In order for the model to converge, one or both of the following restrictions were implemented: *βampS = βampC, βmultS=βmultC*.

The final parameters included in the model are reported in Table (ii).

Model Fit

The sum of squared error (SSE) was used to compare models. The model with lower SSE is preferred. We chose not to use a criterion which penalizes for the number of parameters because we preferred a model with a better fit over one that is more parsimonious.

Table (i) displays the final models along with SSE for the national and provincial data. Overall, based on our criteria, the AR model performed better and we chose to use that model throughout the paper for consistency.

Imputing missing data for KwaZulu-Natal (KZN)

To impute the incidence of microbiologically-confirmed PTB for KZN for 2004 to 2010, the model below (3) was fitted to data from the three provinces bordering KZN for 2004-2012, as well as data for KZN in 2011 and 2012.

$$\begin{aligned} y_t &= \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 prov + \epsilon_t \\ \epsilon_t &= w_t - \phi_1 \epsilon_{t-1} - \ldots - \phi_{12} \epsilon_{t-12} \\ w_t &\sim N(0, \sigma^2) \end{aligned} \tag{3}$$

where the model is as described for (1) and *prov* represents the province: Eastern Cape, Free State, KZN or Mpumalanga.

We compare the estimated rates to those from the ETD, shown in Figure S4. As highlighted in the manuscript, mPTB incidence was consistently higher than PTBr incidence nationally and in seven of nine provinces.

Adjusting for testing

To investigate the influence of testing rate in our model, we fit the model (1) with an added term for the ratio of number of tests per positive case. The estimated incidence rates were similar to those without adjusting for testing rate (Figure (a)). Because there were no drastic differences in the model fit, the additional term to adjust for testing rates was omitted from the final model.

We did not investigate the impact of performance on testing rates. LPA tests were only used from 2009 and culture was not widely available until 2006, in addition these tests represent only a small fraction of the data. Xpert MTB/Rif has a much higher sensitivity than the other tests, but was only introduced in 2011. The impact of the introduction of this test is only believed to be apparent from 2013 onward (as 100% coverage will be achieved then). Although the data comprise of test results from four different tests with different degrees of performance, the majority of the data is from smear results. The impact of the other tests is believed to be too small to have an influence on the data.

Final Model

After estimating the KZN data, we re-fit model (1) to the national data which included model (2) estimates for KZN. Details of the final model are given in Table (ii). We examined the ACF and PACF plots shown in Figure (b) and the residuals appear to be random white noise.

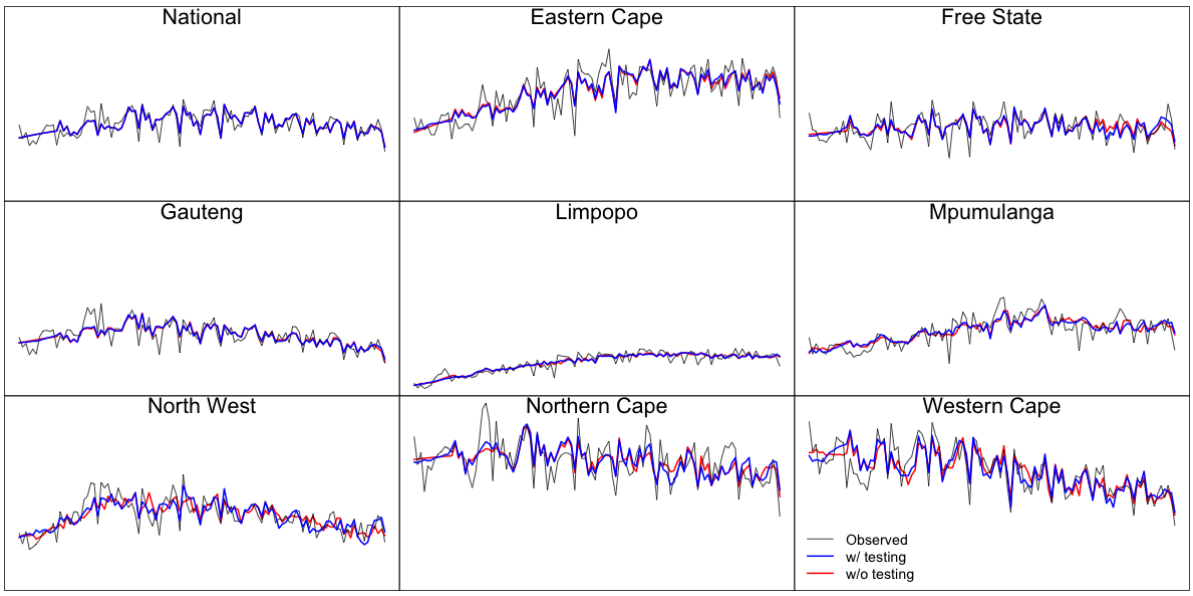

Figure (a): Comparing the estimated fitted values between models with and without adjustment for testing rates

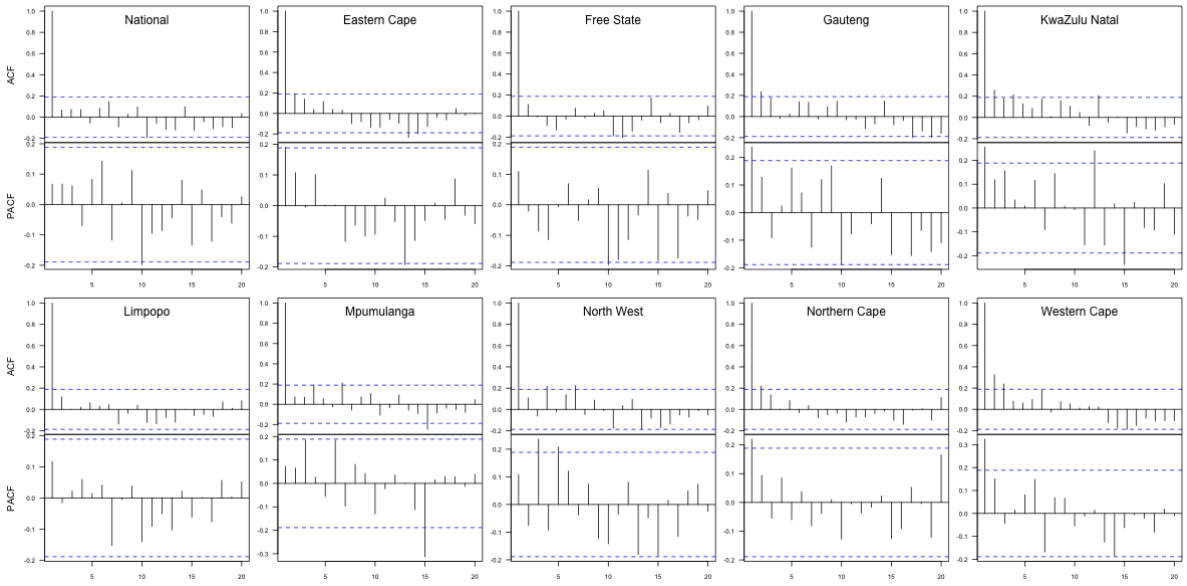

Figure (b): ACF and PCF plots for the final model describing the national and provincial data

Table (i): Comparing AR and NL models. *Final model contains the AR parameters if an AR model and the restrictions if a NL model.

Age group	Region	Model	SSE	Final Model*
All	National	AR	2303	12
		NL	3819	
	EC	AR	7777	12
		NL	6833	
	FS	AR	4555	12
		NL	5451	
	GP	AR	3287	12
		NL	4203	
	LP	AR	1012	2,12
		NL	1007	
	MP	AR	4589	1,2,12
		NL	4998	
	NW	AR	6426	1,2,11,12
		NL	7791	
	NC	AR	13264	12
		NL	14982	
	WC	AR	8111	2,10,12
		NL	11723	

<5	National	AR	58	1,9,12
		NL	49	
	EC	AR	261	1,12
		NL	258	
	FS	AR	116	4,10,12
		NL	113	
	GP	AR	91	1,2,3,5
		NL	113	
	LP	AR	41	1,4
		NL	41	
	MP	AR	141	1,3,5
		NL	158	
	NW	AR	389	1
		NL	494	
	NC	AR	841	1
		NL	993	
	WC	AR	327	1,10
		NL	336	
15-24	National	AR	1375	12
		NL	2124	
	EC	AR	4678	9,12
		NL	4092	
	FS	AR	3101	12
		NL	3180	
	GP	AR	1621	1,12
		NL	2056	
	LP	AR	785	
		NL	693	
	MP	AR	2716	1,12
		NL	2442	
	NW	AR	3701	1,2,11,12
		NL	4807	
	NC	AR	11018	1,6,7,12
		NL	14186	
	WC	AR	9242	12
		NL	12499	
25-44	National	AR	7427	12
		NL	7510	
	EC	AR	35678	12
		NL	40506	
	FS	AR	18417	12
		NL	21992	
	GP	AR	7200	12

		NL	6296	
	LP	AR	3823	2,12
		NL	3422	
	MP	AR	17394	1,2,12
		NL	20273	
	NW	AR	19216	1,3,9,12
		NL	22148	
	NC	AR	47331	12
		NL	37694	
	WC	AR	20032	12
		NL	26297	
45-64	National	AR	5446	12
		NL	8044	
	EC	AR	29238	4,10,12
		NL	33911	
	FS	AR	13780	11,12
		NL	14877	
	GP	AR	4298	12
		NL	3124	
	LP	AR	4221	12
		NL	3987	
	MP	AR	16298	1,2,12
		NL	21247	
	NW	AR	19177	1,2,11,12
		NL	21933	
	NC	AR	41183	1,12
		NL	45149	
	WC	AR	14759	2,10,12
		NL	18869	
>65	National	AR	1502	10,12
		NL	1431	
	EC	AR	10867	1,4,9,12
		NL	13018	
	FS	AR	4462	10
		NL	4104	
	GP	AR	1335	
		NL	1156	
	LP	AR	2163	2
		NL	2238	$\beta_{ampS}=\beta_{ampC},\beta_{multS}=\beta_{multC}$
	MP	AR	4352	1
		NL	4626	
	NW	AR	6501	1,2
		NL	6841	

NC	AR	13342	7,12
	NL	15259	
WC	AR	4096	1,4,12
	NL	3950	

Table (ii): Results from final models

Age Group	Region	Peak	Acceleration	AR Parameters
All	National	Jun-09	0.009506	12
	EC	Nov-10	0.011505	12
	FS	Jan-09	0.004001	12
	GP	Nov-07	0.009927	12
	KZN	Jul-11	0.011492	2,5,10,12
	LP	Nov-10	0.00661	2,12
	MP	Dec-10	0.008579	1,2,12
	NW	Jun-08	0.016424	1,2,11,12
	NC	Nov-06	0.008115	12
	WC	no peak	0.004345	2,10,12
<5	National	Jun-09	0.001468	1,9,12
	EC	Oct-09	0.002159	1,12
	FS	Jun-09	0.001269	4,10,12
	GP	Jul-08	0.001575	1,2,3,5
	KZN	Jul-12	0.000892	1,8,9,12
	LP	Jul-11	0.000617	1,4
	MP	Dec-10	0.001431	1,3,5
	NW	Apr-08	0.003106	1
	NC	May-08	0.001637	1
	WC	May-07	0.001778	1,10
15-24	National	Jan-08	0.005153	12
	EC	Sep-09	0.012047	9,12
	FS	Sep-10	0.000698	12
	GP	Sep-07	0.006861	1,12
	KZN	Apr-11	-0.00268	1,8,12
	LP	Sep-09	0.00561	
	MP	Jul-08	0.008645	1,12
	NW	Oct-07	0.010096	1,2,11,12
	NC	no peak	0.002769	1,6,7,12
	WC	Oct-04	0.005721	12
25-44	National	Jun-08	0.016577	12

	EC	Oct-09	0.032133	12
	FS	Jan-08	0.009822	12
	GP	Mar-08	0.018272	12
	KZN	Dec-07	0.004982	1,2,10,12
	LP	Oct-09	0.017671	2,12
	MP	Sep-08	0.027008	1,2,12
	NW	May-08	0.031382	1,3,9,12
	NC	Apr-06	0.014771	12
	WC	Feb-04	0.007398	12
45-64	National	Oct-08	0.013623	12
	EC	May-10	0.033072	4,10,12
	FS	Oct-09	0.007151	11,12
	GP	Jun-08	0.013204	12
	KZN	Dec-46	-0.00054	1
	LP	Jun-10	0.015175	12
	MP	Feb-09	0.021294	1,2,12
	NW	Aug-08	0.029352	1,2,11,12
	NC	Jul-08	0.018732	1,12
	WC	Apr-05	0.007043	2,10,12
>65	National	Jan-09	0.006074	12
	EC	Jul-10	0.021611	1,4,9,12
	FS	Jan-11	0.002721	10
	GP	Jul-08	0.003881	
	KZN	Mar-12	-0.00657	1,9
	LP	Feb-10	0.007965	2
	MP	Jun-09	0.009563	1
	NW	Sep-08	0.011409	1,2
	NC	Mar-06	0.00562	7,12
	WC	Apr-06	0.005976	1,4,12

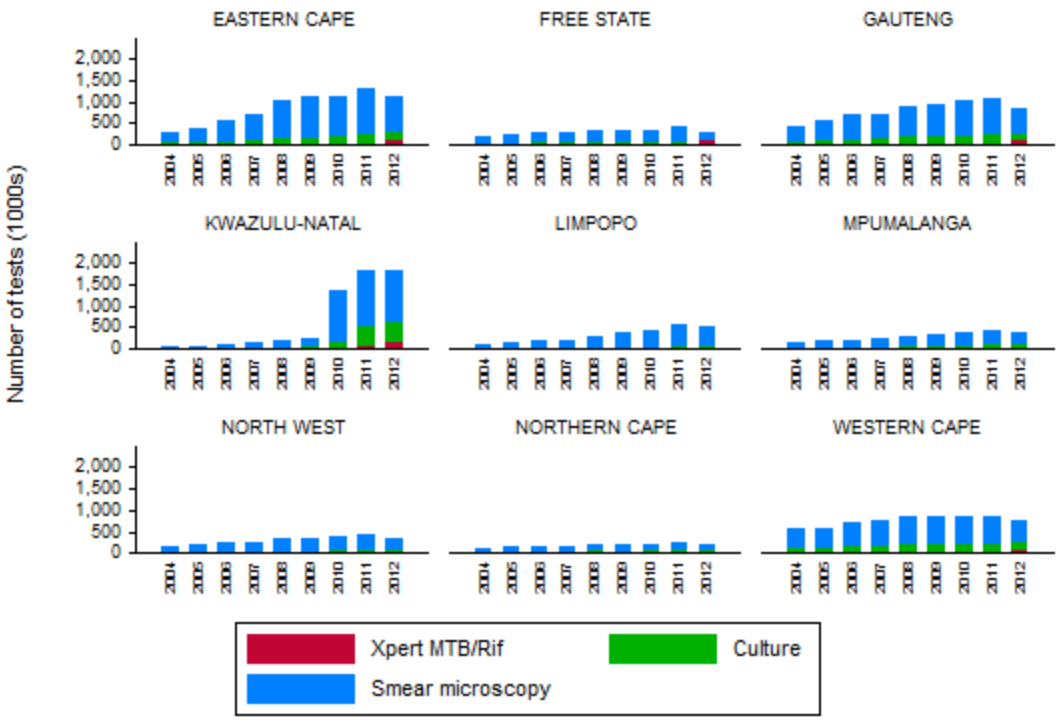

Figure S1: Numbers of tests by diagnostic method, per province in South Africa, 2004-2012*

Table S1: Difference in microbiologically confirmed pulmonary TB case numbers and incident rates when a 12-month interval is used to identify new episodes, compared to a 24-month interval

Province**	2005		2006		2007		2008		2009		2010		2011		2012	
	Δ n††	Δ I‡‡	Δ n	Δ I	Δ n	Δ I	Δ n	Δ I	Δ n	Δ I	Δ n	Δ I	Δ n	Δ I	Δ n	Δ I
Eastern Cape	105	0.25	266	0.43	308	0.52	368	0.52	303	0.45	348	0.46	288	0.36	313	0.45
Free State	136	0.75	221	1.12	271	1.42	250	1.17	214	1.09	179	0.96	165	0.86	150	0.73
Gauteng	148	0.15	259	0.30	346	0.48	386	0.47	391	0.51	415	0.52	370	0.55	333	0.64
Limpopo	22	0.00	68	0.33	66	0.28	89	0.49	106	0.47	110	0.49	96	0.49	41	0.24
Mpumalanga	38	0.21	67	0.37	106	0.48	116	0.41	173	0.52	134	0.46	140	0.44	50	0.16
North West	93	0.39	215	0.83	247	0.90	219	0.88	205	0.82	235	0.99	214	0.96	72	0.33
Northern Cape	122	0.86	174	1.24	173	1.36	174	1.32	209	1.59	157	1.22	127	0.93	107	0.89
Western Cape	462	0.79	915	1.35	920	1.38	1006	1.56	1101	1.90	1095	1.88	1078	1.97	1136	2.10
South Africa	1,126	0.43	2,185	0.78	2,437	0.79	2,608	0.87	2,702	0.91	2,673	0.94	2,478	0.83	2,202	0.74

* Prior to 2006, there was limited availability of facilities to undertake culture for *Mycobacterium tuberculosis* in some settings, and utilization of this diagnostic method may not have been optimal. Xpert MTB/Rif was introduced in South Africa in 2011 and the roll-out of this new molecular diagnostic tool was only completed in October 2013.

** KwaZulu-Natal was excluded from the sensitivity analysis because of the incomplete data coverage for 2004 to 2010

†† **Δ n** is the difference in number of mPTB cases when using a 12-month interval to identify new cases compared to a 24-month interval

‡‡ **Δ I** is the difference in TB incidence rates/100000 when using a 12-month interval to identify new cases compared to a 24-month interval

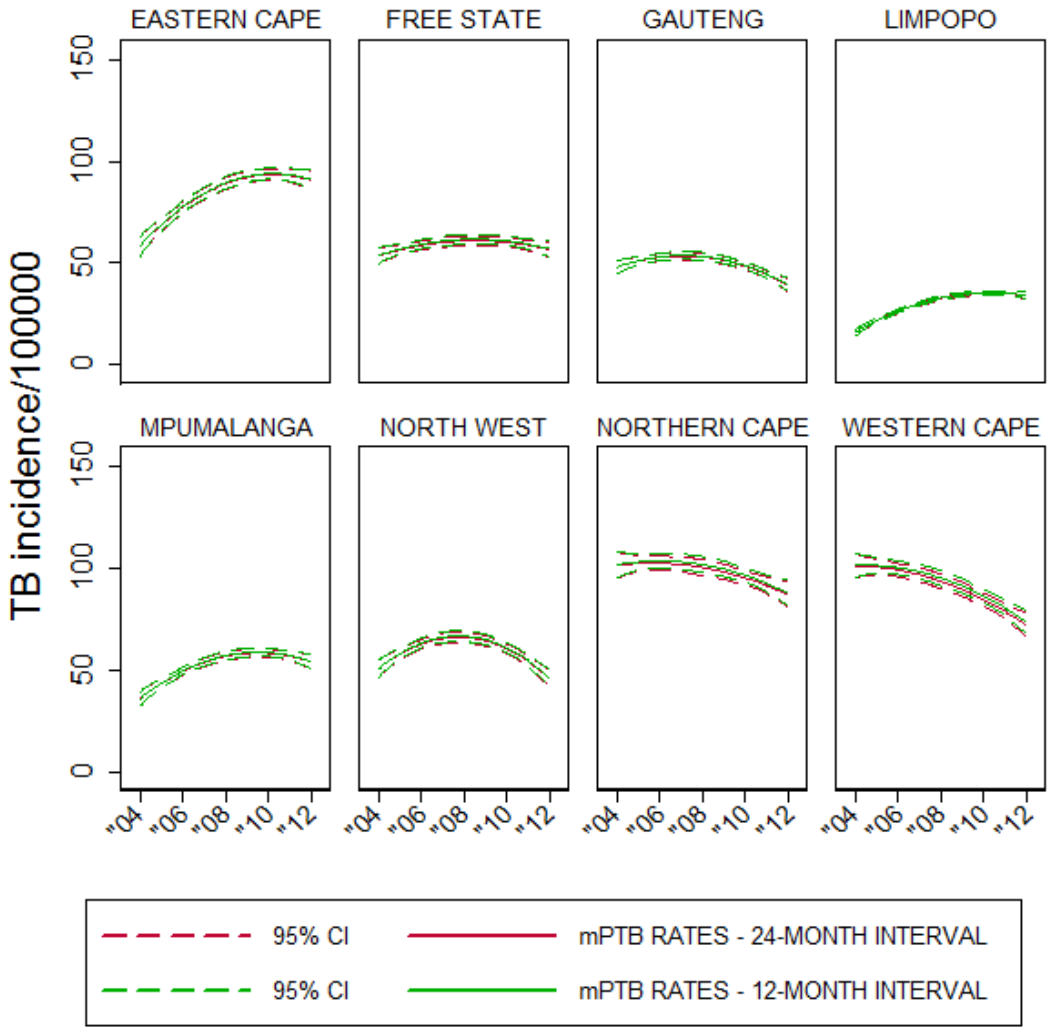

Figure S2: Comparison of microbiologically confirmed pulmonary TB incidence rates/100000 population using a 12-month and a 24-month interval for identification of new episodes

Table S2: Number of people tested for tuberculosis per 100,000 population, by province in South Africa, 2004-2012

Province	2004	2005	2006	2007	2008	2009	2010	2011	2012
Eastern Cape	2,779	3,609	4,889	5,978	8,763	9,155	9,193	10,799	9,756
Free State	3,663	3,917	4,450	4,638	5,215	5,109	5,791	6,859	6,097
Gauteng	2,452	3,123	3,522	3,330	3,858	3,905	4,289	4,432	3,781
Kwazulu-Natal[§§]	-	-	-	-	-	-	-	10,240	9,653
Limpopo	1,165	1,571	2,016	2,360	3,284	3,875	4,275	6,240	5,245
Mpumalanga	2,536	2,574	3,115	3,282	3,823	4,175	4,150	4,920	4,200
North West	2,639	3,406	3,833	4,103	5,013	5,017	5,760	6,137	5,369
Northern Cape	5,042	5,755	6,546	6,385	7,179	7,100	8,824	9,908	8,223
Western Cape	5,440	5,495	6,195	6,154	6,506	6,018	5,762	6,013	5,607
Nationally (including limited KZN data 2004-2010)	2,353	2,773	3,347	3,559	4,376	4,559	6,160	7,164	6,415
Nationally (Excluding KZN)	2,926	3,423	4,053	4,264	5,228	5,333	5,615	6,403	5,616

[§§] Limited data from the sole culture site is available for KZN between 2004 and 2010.

Table S3: Relative increases in antiretroviral treatment uptake in South Africa, 2004-2012

Time period	**2004-2005**	**2005-2006**	**2006-2007**	**2007-2008**	**2008-2009**	**2009-2010**	**2010-2011**	**2011-2012**
Difference in ART coverage between consecutive years (slope)	1·19	2·25	2·61	3·58	4·92	4·57	4·08	3·59

	2004-05 and 2005-06	**2005-06 and 2006-07**	**2006-07 and 2007-08**	**2007-08 and 2008-09**	**2008-09 and 2009-10**	**2009-10 and 2010-11**	**2010-11 and 2011-12**
Change in slope between consecutive two-year time periods	**1·05**	0·36	0·96	**1·34**	-0·35	-0·49	-0·49

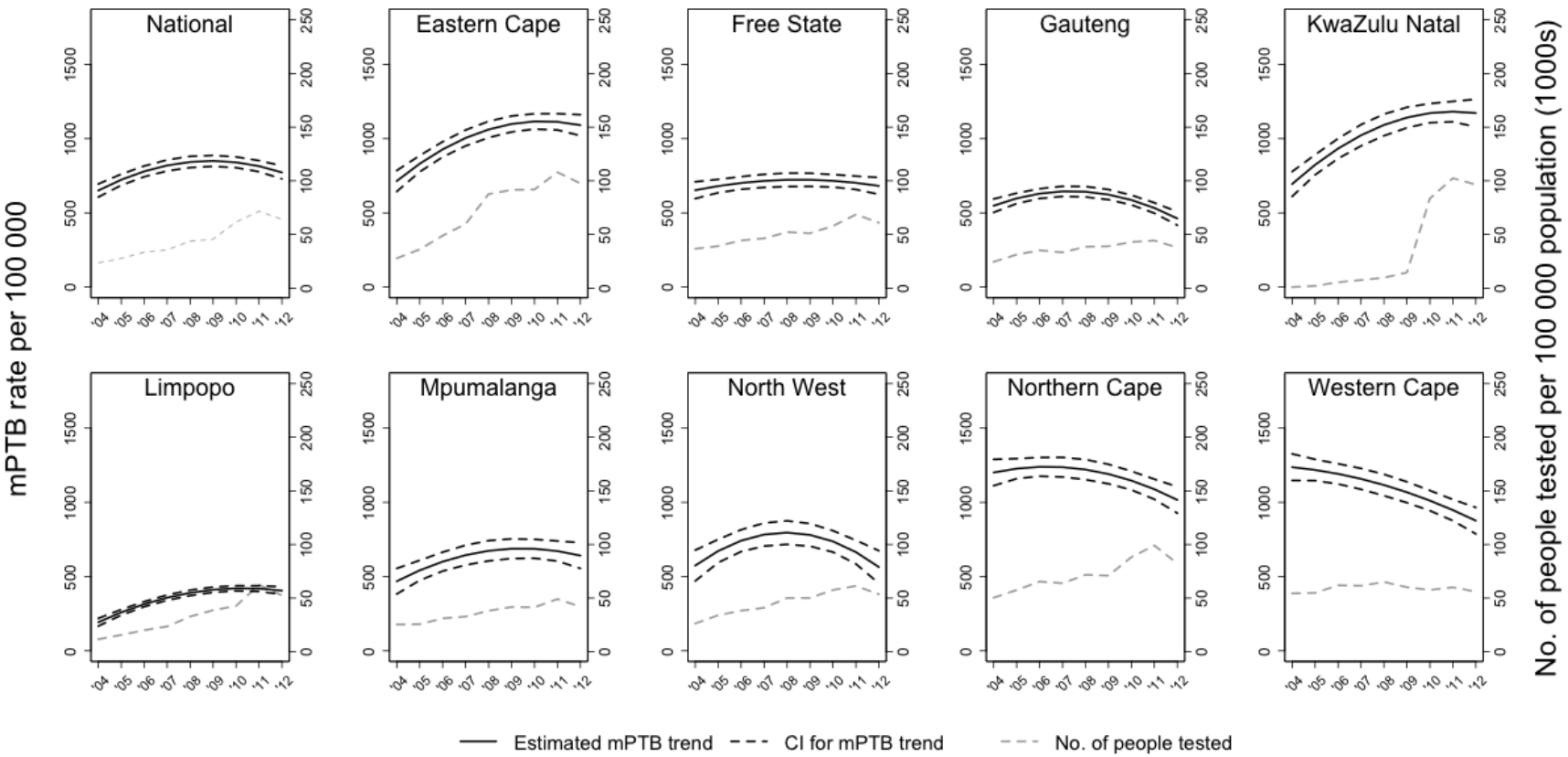

Figure S3: Incidence of microbiologically-confirmed pulmonary tuberculosis and testing rates (per 100,000 population) in South Africa, nationally and by province in 2004 to 2012

Table S4: Number of tests (smear, culture and Xpert MTB/RIF) done per 100,000 population, by province in South Africa, 2004-2012

Province	2004	2005	2006	2007	2008	2009	2010	2011	2012
Eastern Cape	4,629	6,099	8,619	10,801	16,013	16,934	17,497	20,201	16,735
Free State	7,284	8,158	9,521	10,381	11,504	11,462	12,427	14,721	10,246
Gauteng	4,044	5,349	6,271	6,359	7,916	7,867	8,541	8,753	6,908
Kwazulu-Natal[***]	189	386	1,045	1,403	1,769	2,403	13,587	17,856	17,541
Limpopo	1,931	2,704	3,579	4,248	5,893	7,106	7,984	10,901	9,192
Mpumalanga	4,333	4,701	5,769	6,305	7,937	9,101	9,240	10,745	8,859
North West	4,667	6,227	7,493	8,172	10,088	10,228	12,014	12,278	9,574
Northern Cape	10,576	12,723	15,030	15,157	17,234	17,054	19,302	20,678	16,817
Western Cape	11,487	11,601	13,389	14,404	15,583	15,647	14,922	14,728	12,959
SA	4,321	5,183	6,448	7,150	8,931	9,390	12,086	13,886	11,899

Table S5: Comparison of number of pulmonary tuberculosis cases registered by the National Tuberculosis Control Program of South Africa and microbiologically-confirmed pulmonary TB cases identified through the National Health Laboratory Service Corporate Data Warehouse.

	2004	2005	2006	2007	2008	2009	2010	2011	2012
Eastern Cape	25	29	33	25	23	20	21	25	29
Free State	15	5	8	-4	-1	-1	-1	-6	2
Gauteng	48	53	53		42	34	[†††]	30	27
Kwazulu-Natal	45	53	56	34	26	9		26	28
Limpopo	17	26	26	23	18	15	17	17	26
Mpumalanga					28	21		20	30
North West	6	15	21	14	11	-8	-11	-16	-3
Northern Cape	44	41					24	29	
Western Cape	37	28	35	34	27	21	20	19	20
South Africa	**30**	**31**	**33**	**32**	**30**	**14**	**11**	**16**	**20**

[***] Between 2004 and 2010, there was limited data available for KZN from the sole culture lab in the province, from 2011 onwards electronic data capture was implemented across the province

[†††] Shaded blocks indicate years and provinces for which there was limited/no data from ETD

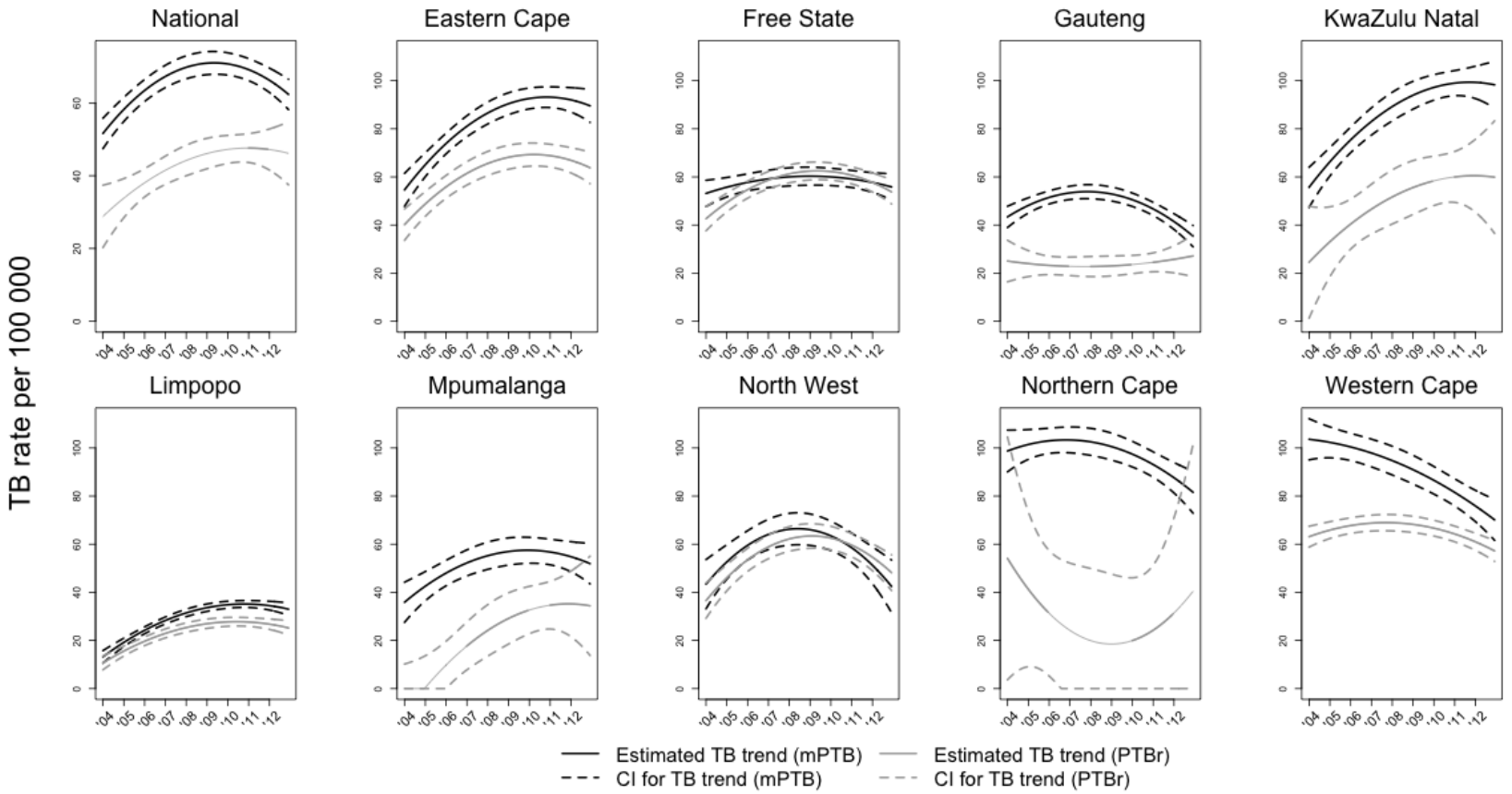

Figure S4: Trends in microbiologically-confirmed pulmonary TB (mPTB) and registered pulmonary TB (PTBr) rates, nationally and per province in South Africa, 2004-2012

Feint lines indicated years for which data on PTBr cases where incomplete or missing.

References

1. Statistics South Africa. Use of health facilities and levels of selected health conditions in South Africa : Findings from the General Household Survey, 2011. Pretoria, 2013.
2. National Department of Health. The National Antenatal Sentinel HIV and Syphilis Prevalence Survey in South Africa. Pretoria, 2012.
3. Lutge E, Lewin S, Volmink J, Friedman I, Lombard C. Economic support to improve tuberculosis treatment outcomes in South Africa : a pragmatic cluster-randomized controlled trial. *Trials* 2013; **14**: 1.