Global Epidemic Trend of Tuberculosis during 1990-2010: Using Segmented Regression Model

Anoushiravan Kazemnejad (PhD)\textsuperscript{a}, Shahram Arsang Jang (MSc)\textsuperscript{b,c}, Firouz Amani (PhD)\textsuperscript{a}, Alireza Omidi (MSc)\textsuperscript{d}

\textsuperscript{a} Department of Biostatistics, Faculty of Medical Sciences, Tabriz Modares University, Tehran, Iran
\textsuperscript{b} Department of Epidemiology and Biostatistics, Faculty of Health, Qom University of Medical Sciences, Qom, Iran
\textsuperscript{c} Department of Biostatistics, Faculty of Health, Ardabil University of Medical Sciences, Ardabil, Iran
\textsuperscript{d} Department of Public Health, Faculty of Health, Qom University of Medical Sciences, Qom, Iran

\section*{ABSTRACT}

\textbf{Background:} Tuberculosis (TB) is a pandemic disease. It is the second leading cause of death from infectious diseases after human immunodeficiency virus (HIV) in the world. The main objective of this paper was to determine and compare the epidemiology of TB incidence rate and its trend changes during 1990-2010 in six WHO regions regarding age, gender and income levels.

\textbf{Methods:} The Average Annual Percent Change (AAPC) and Annual Percent Change (APC) of TB incidence, mortality, treatment-successes, case detection rates, as well as change points of trend was estimated using segmented regression model. The number of change points was selected by the permutation procedure based on likelihood ratio test.

\textbf{Results:} Two change points for global TB incidence rate trend with AAPC\textsubscript{years} equaling -1.4 \% was estimated, the maximum AAPC\textsubscript{years} of six regions was attributed to the American region (-3.5\%). AAPC of TB treatment-successes rate for Eastern Mediterranean (+2.2\%), the Americas (+1.6\%), south East Asia (+8\%) and Global (+1.1\%) were significant (P<0.05). Moreover AAPC\textsubscript{years} of TB case detection rate for South East Asia (+7.5\%), Eastern Mediterranean (+4.9\%), Africa (+2.8\%) and the Americas (+1.7\%) were significant (P<0.05). Globally, all of income categories had descending trend of TB incidence and mortality rate, except the upper-middle income level that had ascending incidence trend (AAPC=+0.7\%).

\textbf{Conclusions:} Globally, TB incidence and mortality rates have downturn trend and TB treatment successes and detection rates have upward trend, but their changes rate are insufficient to reach the goal of TB stop strategy. The economic levels have effect on trend, with no clear pattern, so it seems necessary that evaluation TB control programs based on characteristics of countries for reach TB control goals.

\section*{Introduction}

Tuberculosis (TB) is an infectious bacterial disease caused by \textit{Mycobacterium tuberculosis}. It is the second leading cause of death worldwide amongst communicable diseases, striking 8.8 million people and killing 1.4 million in 2010\textsuperscript{1}.

Characterizing the trend of diseases and its changes synchronically can have an important role for evaluating the success of disease control strategies, health development indicators and health planning. Segmented linear regression is also known as breakpoint linear regression- piecewise linear regression- used for analysis and trend description of continuous data from mortality or incidence rate\textsuperscript{2,3}. Understanding points of changes and number of changes are important. Permutation procedure, Bayesian Information Criteria (BIC), Akaike Information Criterion (AIC), and Generalized Cross Validation (GCV) can be used to select the number of change points\textsuperscript{4}. Moreover, we can estimate the Annual Percent Changes (AAPC) and Average Annual Percent Changes (AAPC) using these methods. AAPC or APC provide summery statistics of trend. The estimate of change points in segmented regression is based on maximum likelihood, least square and Bayesian analysis. Segmented regression models fit with Lerman’s grid search (LGS) or Hudson’s continuous fitting algorithm (HCA) approaches\textsuperscript{5,6}. Several authors have examined segmented regression. Kim et al. compared cancer trend rate between two groups with segmented regression model and estimated the \textit{P}-value using permutation test\textsuperscript{7}. Arsang et al used the BIC and Joint-point methods in order to determine the epidemiology of tuberculosis in Iran during 2001-2008\textsuperscript{8}; Nemes et al have used this method for detecting the relationship between abnormally expressed genes due to aberrant DNA copy numbers and subsequent altered gene expression profiles\textsuperscript{9}. Wijlaars et al have done the same to assess the effect of supervision of committee on safety of medicines on antidepressant prescription rates changes\textsuperscript{10}; Bouadma et al has
used the above-mentioned model to detect the impact of prevention program on ventilator associated pneumonia. Muggeo et al. have used this model for fitting the trend of fertility data.

The aims of this study were (a) estimating the number, location or time of change points for TB incidence, mortality, SP (Smear Positive) treatment-successes and case detection rate from 1990 to 2010 globally and six WHO regions (211 countries) including Eastern Mediterranean, Africa, The Americas, Europe, Western Pacific, South East Asia; (b) Pair wise comparison of parallelism and coincidence of trend between regions with the economic, age and gender categories as independent variables; (c) Providing the summery statistics of trend for each of segments for the last 5 and 10 years. The present research by determines TB trend changes can help to detect effective factors in TB control, improve current planning polices attributed to TB and prioritization.

**Methods**

In order to determine the TB trend and the number of change points, the segmented regression model was used with maximum three numbers of change points for n=20, in which n is the number of data points (1990 - 2010) as an independent variable. The maximum number of change points depends on the number of data points. TB incidence rate per 10^5 populations, mortality incidence per 10^5 populations, percentage of SP treatment-successes and case detection rate of six WHO regions were considered response variables. Logarithmic transformation on response variable was performed. Parallelism of two combinations was tested in order to find out whether the two (trend of TB among regions were) regression mean functions were parallel. To do so, the years, six WHO regions, country income levels (low, lower-middle and upper-middle income countries), age groups and gender were inserted as independent variables. We used World Bank’s criterion that categorizes countries based on gross national income (GNI) per capita. The groups are: low income, $1,035 or less; lower middle income, $1,036 - $4,085; upper middle income, $4,086 or more. Least squares methods for estimating regression parameters and Grid Search methods to determine the best fit for each individual model were used.

We performed all the statistical analyses using Joinpoint software, V3.5.1 which is available at American Cancer Center. Permutation test and BIC were also used for selecting the known number of change points. Our data on global trends in TB came from WHO collects by WHO regional offices.

**Segmented Regression**

Segmented linear regression is one of the linear regression models used for segmenting nonlinear regression models into linear segments and points among segments called change points. For each segment there is a different \( f_i(x) \) function, describing segmented regression curve for \( r \) segments as:

\[
 f(x) = E[y|x] = \begin{cases} 
 f_1(x; \beta_1) & x \leq \tau_1 \\
 f_2(x; \beta_2) & \tau_1 < x < \tau_2 \\
 \cdots & \cdots \\
 f_r(x; \beta_r) & \tau_{r-1} < x
\end{cases}
\]  

In which \( \tau \) indicates the change point and \( f_i(x, \beta_i) \) regression function of \( i \)th segment. Regarding \( (x_{ij}, y_{ij}) \) pair of the \( j \)th observations for \( i \)th groups \( \{i=1, 2; j=1, \ldots, n_i\} \), e.g. \( y_{ij} \) can be TB incidence rate at the \( j \)th years, \( x_{ij} \) years, for the regional or group \( i \), then regression response function mean for each segment or group \( i \) equals, \( E(y|x) = \beta_0 + \beta_1 x + \delta_{i,1}(x - \tau_{1,1}) + \cdots + \delta_{i,k_i}(x - \tau_{k_i,k_i})^+ \)

Where \( k_i \) is the unknown number of change points, \( \tau_{i,j} \) \( (i=1\ldots k) \) is the unknown change points, \( \beta_{i,j} \) \( (i=1,2) \) and \( \delta_i \) are the regression parameters and \( (x - \tau_{i,j})^+ = (x - \tau_{i,j}) \) for \( (x - \tau_{i,j}) > 0 \).

The Lerman’s grid search was used on \( \tau_{i,1}, \ldots, \tau_{i,k_i} \) that uses the least square estimation method, carrying out a grid search over trend to map minimum value of residuals sum of square function. Parallelism test was used to find out whether two segmented mean regression functions are parallel, and to see whether two segmented mean regression functions are identical, coincident test was used accordingly. Permutation test was used for selection of models with change points against j change points (i<j), the model with i change point was selected if

\[
RSS(i) \leq (1 + c_n(i, j; \infty))RSS(i) = \delta_i^2 h_n(i, j)
\]

In which RSS denotes residual sum of squares, \( c_n \) is a critical value obtained under the null model with i change points \( c_n \) equal the p-value, \( \delta_i^2 = \frac{RSS(M_n)}{(n - M_n)} \) and \( h_n(i, j, \delta) \) is a non-decreasing penalty function for obtaining BIC.

F statistic was used for testing \( H_0 \): there are \( K_0 \) change points against \( H_1 \): there are \( K_M \) change points, in which \( M \) is the maximum number of change points that is determined, \emph{a priori} [by the researcher]. If \( H_0 \) is rejected, then \( H_0 : K_0=1 \) against \( H_1: K_0=M \) will be tested, otherwise \( H_0: K_0=0 \) against \( H_1: K_0=M-1 \) will be tested instead. And this procedure will be continued until hypothesis \( k=i+1, 0<i<M \) and the number of change point denoted by \( \hat{k} \) is estimated.

**APC and AAPC**

Annual Percent change is a way for describing the trend of incidence rate over time, with the assumption that incidence rates at year \( y \) with steady rate change into last \( y-1 \) year. To estimate the APC for a series of data, the following regression model is used:

\[
\log \left( \frac{R_{y+1}}{R_y} \right) = b_0 + b_1 y \quad \text{where} \quad \log \left( R_y \right) = \text{the natural log of the rate in year} \ y. \text{ The APC from year} \ y \text{ to year} \ y+1 = \frac{\left[ e^{b_0+b_1 y+1} - e^{b_0+b_1 y} \right]}{e^{b_0+b_1 y}} \times 100 = (e^{b_1} - 1) \times 100
\]

In order to present the summary of trend and determine the interval of years, AAPC can be used, providing single value of trend for the last 5 and 10 years, even if there are changes throughout the trend. This value is the average weight of APC segmented regression in which the weight equals the length of APC interval. To estimate the AAPC following equation is used:
\[ AAPC = \left\{ \left\{ \exp \left( \frac{\sum w_i b_i}{\sum w_i} \right) - 1 \right\} \times 100 \right\} \]  

That \( b_i \)s are the slope coefficients for each segment in the desired range of years, and the \( w_i \)s are the length of each segment run in the range of years. In the Joinpoint software, the AAPC confidence interval is based on the normal distribution, and the AP confidence interval is based on a t distribution. If an AAPC lies entirely within a single Joinpoint segment, the AAPC is equal to the APC for that segment\(^1,8,10\).

### Results

During 1995-2010 there were 26,815,677 SP case notifications (64.3% males and 35.6% females). The maximum TB incidence rate was attributed to African region in 2004 with 303.36 per 10^5 populations. The maximum mortality rate was attributed to Western Pacific region in 1997 with 44.09 per 10^5 populations. TB incidence and mortality rate (exclude AIDS) globally for the last segment have declined, and two change points for incidence and one change point for mortality rate were estimated throughout the trend, so that the estimated mean segmented regression function for global TB incidence and mortality rate are:

\[ E(y|x) = 12.67 - 0.0038x + 0.006(x - 1996)^+ - 0.016(x - 2003)^+ \] (Incidence rate)

\[ E(y|x) = 32.54 - 0.014x - 0.032(x - 2003)^+ \] (Mortality rate)

TB incidence rate trend in all of the WHO regions for the last segment have declined (AAPC<0). Table 1 shows number and locale of change point(s). The maximum estimated AAPC of TB incidence rate for the last 5 years is attributed to the Americas region (AAPC=3.5) and the minimum value is attributed to Eastern Mediterranean region (AAPC=8). The AAPC result of TB incidence and mortality incidence rate for the last 5 and 10 years are shown in Table 2.

#### Table 1: Number and local of change points in tuberculosis (TB) incidence, mortality, case detection, and smear positive (SP) treatment-success rate trend

<table>
<thead>
<tr>
<th>WHO regions</th>
<th>TB Incidence Rate</th>
<th>TB Mortality Incidence Rate</th>
<th>TB SP treatment-success rate</th>
<th>TB case detection rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Americas</td>
<td>0</td>
<td>1(1996)</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>South East Asia</td>
<td>1(2004)</td>
<td>0</td>
<td>0</td>
<td>1(1995)</td>
</tr>
</tbody>
</table>

Year : downward shape after change point; Year : upward shape after change point; Year : fixed shape after change point

#### Table 2: The Average Annual Percent Changes (AAPC) for tuberculosis (TB) incidence, mortality case detection, and smear positive (SP) treatment success rate for last 5 and 10 years

<table>
<thead>
<tr>
<th>WHO regions</th>
<th>AAPC TB incidence rate(^a)</th>
<th>AAPC TB Mortality incidence rate *</th>
<th>AAPC of Smear-positive TB treatment-success rate</th>
<th>AAPC of TB case detection rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern Mediterranean</td>
<td>-0.8  -0.8</td>
<td>-6.0  -5.2</td>
<td>2.2  2.2</td>
<td>4.9  4.9</td>
</tr>
<tr>
<td>Africa</td>
<td>-0.4  -1.6</td>
<td>-2.2  -3.1</td>
<td>0.1  0.1</td>
<td>2.8  2.8</td>
</tr>
<tr>
<td>The Americas</td>
<td>-3.5  -3.5</td>
<td>-5.2  -5.2</td>
<td>1.6  1.6</td>
<td>1.7  1.7</td>
</tr>
<tr>
<td>Europe</td>
<td>-1.8  -1.8</td>
<td>-3.1  -3.9</td>
<td>0.5  0.5</td>
<td>-0.8 -2.1</td>
</tr>
<tr>
<td>Western Pacific</td>
<td>-2.6  -2.6</td>
<td>-4.6  -5.4</td>
<td>6.9^a  2.0</td>
<td>7.5  7.5</td>
</tr>
<tr>
<td>South East Asia</td>
<td>-1.1  -1.7</td>
<td>-5.1  -5.1</td>
<td>0.8^a  0.8^a</td>
<td>7.5  7.5</td>
</tr>
<tr>
<td>Global</td>
<td>-1.0  -1.4</td>
<td>-3.9  -4.6</td>
<td>2.7^a  1.1^a</td>
<td>4.3  2.0</td>
</tr>
</tbody>
</table>

\(^a\)AAPC is significantly different from zero at \(P<0.05\)

While the global TB incidence rate during 1990-1996 has declined (AAPC= -0.008), it has been constant during 1996-2004 (AAPC=0.2, \(P=0.086\)) and has declined for the last segment (AAPC= -1.4, \(P<0.001\)).

The parallelism hypotheses for TB incidence as well as mortality trend between WHO regions were rejected (\(P>0.05\)). The results of APC and estimated trends for TB mortality rate are shown in Figure 1.

The number and locale of change points of TB incidence, mortality, case detection, and SP treatment-success rate for WHO regions are shown in Table 1.

During 1999-2010 the maximum average value of treatment-success was attributed to South East Asia (89.07%) and the maximum average value of case detection rate was attributed to Europe (70.4%). The APC of TB SP treatment-success rates relating to the last segment were for Eastern Mediterranean (AAPC=2.2, \(P<0.001\)), the Americas (AAPC=+1.6, \(P=0.007\)), South East Asia (AAPC=0.8, \(P=0.016\)) and global (AAPC=1.1, \(P<0.001\)). The APC of TB case detection rates relating to the last segment were for Eastern Mediterranean (AAPC=+4.9, \(P<0.001\)), Africa (AAPC=+2.8, \(P<0.001\)), the Americas (AAPC=+1.7, \(P=0.001\)) and South East Asia (AAPC=+7.5, \(P=0.003\)). The TB case detection rate trend between the Americas and Europe was parallel (\(P=0.180\)) and coincident (\(P=0.171\)). The parallelism and coincident hypotheses between other WHO regions were rejected (\(P<0.05\)).

Trend of TB incidence as well as mortality rate were analyzed separately based on country income categories, gender and age groups. AAPC results of TB incidence rate and mortality incidence rate for the last five years of low, lower-middle and upper middle income are shown in Figure 2 and 3.

Pair-wise comparison demonstrates that AAPC5years differences of TB incidence rate between three segments (income level) for all regions were significant at 0.05 level, with the exclusion of the Americas (\(P>0.1\)), the low and lower-middle income countries globally and the low and lower-middle income countries of Europe. The global TB incidence has changed in 2003 from ascending to downturn trend for low (AAPC= -1.3, 95% CI\(_{AAPC} : -1.7, -1\)) and lower-middle income countries (AAPC= -1.5, 95% CI\(_{AAPC} : -1.3, -1\)) but only its upward slopes is reduced for upper-middle income countries (AAPC= -0.69, 95% CI\(_{AAPC} : 0.3, 1.1\)).
Global Epidemic Trend of Tuberculosis

The global TB mortality trends of income categories were downturn for all segments and their slopes are increased after each change point.

We rejected the hypothesis that AAPC 5 years of TB incidence rate is different from zero for Eastern Mediterranean with low, Africa with upper and lower-middle, Europe with low and lower-middle and South East Asia with lower-middle income countries. The AAPC 5 years of other regions and categories were different from zero ($P<0.05$) (Figure 2).

During 1990-2010, the average proportion of TB notified in males in all regions for all age groups, except 0-14 age groups, was 2.018 times higher than females. With the passage of time (1990 to 2010) notification rate trend gap between female and male has been higher than that at previous times. Maximum TB notified during 1990-2010 was attributed to 25-44 age groups in male and 15-34 age groups in females. Globally, the smallest and the largest AAPC 5 years differences between males and females were 1.2 (for age 0-
Anoushiravan Kazemnejad et al.

We rejected the hypothesis that the AAPC of TB notified between two gender groups of all age groups for all regions are different, except some age groups of Eastern Mediterranean. In this region, AAPC differences of TB notified between males and females for 0-14 age groups (-6.2%; 95% CI: -9.4, -2.3), 15-24 age groups (-3.2%; 95% CI: -4.4, -1.3) and 25-40 age groups (-3.5; 95% CI: -5.1, -1.1) were significant, with AAPC for females being higher than males (P<0.001).

The output of BIC and permutation test for selection between ordinary or segmented regression and selection number of change points have produced similar results.

**Figure 2:** Average annual percent change (AAPC) of tuberculosis (TB) incidence rate for last 5 years. (NS: non-significant)

**Figure 3:** Average annual percent change (AAPC) of TB mortality rate for last 5 years. (NS: non-significant)

**Discussion**

The analysis of trend clearly indicates that compared to global trend, incidence reduction rate for all of the regions except Eastern Mediterranean (AAPC>-1.4), and mortality rate reduction for all of the regions except Europe and Africa were higher than those of global trend (AAPC>-4). In addition, mortality rate changes and decline were faster than TB incidence rate.

It can be said that the maximum rate of case detection for the last 5 and 10 years are attributed to South East Asia and the maximum rate of SP treatment-successes for the last 5 and 10 years to Eastern Mediterranean and Western Pacific regions. Since 2006 for the Americas and Western Pacific, 1996 for South East Asia and 2007 for globally, case detection has reached the goal of 85% TB SP treatment-success rate. AAPC of case detection rate during 2005-2010 for Europe and Western Pacific, also SP treatment successes rate for Africa, Europe and Western Pacific are not significant (fixed). Therefore, it seems necessary that attention to and re-evaluation of TB control programs be done for regions with fixed trend that have not reached the TB control goals (Africa, Europe and Eastern Mediterranean).

According to the findings: 1- TB incidence and mortality incidence rate of the WHO regions for the last segment have descent trend with different rates. 2- As regards the fact that more countries (a total of 180) implemented DOTS strategies from 1995 to 2003, and that change points occurred in
2003 (fix trend to downward) globally and in 2004 (decreasing incidence rate) for all low-income countries, it can be concluded that TB control programs have been successful\(^{17-18}\), but there is still a long way for the strategies to reach the goals of TB controls. If TB control programs continue with this trend and rate, it is difficult to reach the goal of stop TB and it is hoped that by 2050 only Americas and Western Pacific regions will reach the goal of incidence rate below one per 10\(^3\) populations.

Like similar studies\(^9\), uniform effects of TB control programs would not be detected in this analysis, because social, biological and TB control program implementation variables are attributed to trend. TB incidence and mortality rate decline in most country were not remarkable; Only the American region has enjoyed the rate of decline (AAPC>4%), although the decline rate in some of them is the same but with regard to TB incidence rate for this region it is almost difficult to reach the goal of TB control at the same time. Therefore, it is vital, in the first step, to consider regions with high incidence and low change rate, and reinforce the DOTS strategy.

Upward trend in TB incidence rate and downward trend of mortality rate globally for upper-middle income countries can be the result of improving the TB detection case that masked decline in TB incidence\(^20\). Moreover, upward trend in TB incidence and mortality rate for upper-middle Eastern Mediterranean can be as a detection case after long delays (regarding the fact that AAPC of mortality is greater than incidence). Unlike previous studies\(^7\), fixed trend of incidence and upward mortality trend for low-income countries of Europe can be the result of decline or fixity of TB treatment successes, increase in extra pulmonary TB and multi-drug resistant\(^9\).

As in other studies\(^7\), there were low TB incidence in 0-14 age groups and high incidence in 15-44 age groups. It may be that exposure with TB risk factors are higher in younger people, compared to other age groups\(^21\). The ratio of TB cases notified in male were higher than females (except 0-14 age groups), which are expected to result from biological, behavioral and sociocultural components differences in male and female susceptibility to \(M.\) \(tuberculosis\) infection or the development of TB disease\(^22\).

Then incidence it seems that TB control strategy could deal with economic situation, so that regions with low income sometime have higher rate of reducing the TB incidence or mortality. This can be regarded as one of the successes in controlling TB, but not enough of TB stop strategies, because there are factors such as immigration, nutrition, early case detection, and access to health care service that impact the TB control programs. So to remove TB, attention must be paid to all forms of TB and the treatment processes should be improved (particularly early case detection).

Permutation test and BIC results demonstrate that when the relation between response and explanatory variable is nonlinear and there is/are change point(s) throughout trend, segmented regression provides unbiased estimation rather than nonlinear regression or Poisson models \(^{13,19}\). Segmented regression can be useful to detect the number and location of changes and also for comparing the trend of disease. Besides, this method can be used for dose-response study to detect the maximum effectiveness dose of drug and compare the parallelism and trend between drugs or groups. In this study we demonstrated TB trend for WHO regions, so for more details about trend, it is indispensable to determine the epidemiology of TB and effective factors such as comorbidities, migration, drug resistance and severity of TB program implementation on trend changes, can be studied inside the regions.

**Conclusions**

As a result, the trend analyses provided useful measurement for comparing the successes the TB control strategies over time among the groups. These findings demonstrated that during 1990-2010, globally TB incidence and mortality rates have downturn trend. In addition, TB treatment successes and detection rates have upward trend, but there changes rate are insufficient to reach the goal of TB stop strategy. The economic levels have effect on AAPC of TB incidence and mortality rates, with no clear pattern, so it seems necessary that evaluation TB control programs based on characteristics of countries for reach TB control goals. The regions that have fixed or negative AAPC for TB incidence rate trend and positive AAPC of TB mortality rate trend, and also young age groups (15-44 years) need to more attention. As results, the America is in better situation than other regions.

**Acknowledgments**

The authors would like to thank the Department of Biostatistics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran and all the study participants for their tremendous cooperation and support.

**Conflict of interest statement**

The authors have nothing to declare.

**Funding**

This study was funded by Tarbiat Modares University.

**References**


