Scientific Research

Prediction of the Number of Tuberculosis Cases and Estimation of Its Treatment Cost in Saudi Arabia Using Proxy Information

Mohamed M. Shoukri[1]*, Bright Varghese[2], Sahal Al-Hajoj[2], Futwan Al-Mohanna[3]

[1]National Biotechnology Centre, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia
[2]Mycobacteriology Research Section, Department of Infection and Immunity, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia
[3]Department of Cell Biology, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia
Email: *shoukri@kfshrc.edu.sa

Abstract

It is well known that the number of people with Tuberculosis (TB) and those who develop multi-drug resistance (MDR) are the fundamental components that affect the total cost of treatment of TB. This paper has two-fold objectives. Firstly, we use the Generalized Linear Regression Models (GLM) to predict the future count of persons with TB and MDR. Due to the fact that assessment of TB cost is methodologically difficult, and compounded with the lack of concrete information about the treatment cost in Saudi Arabia, our second objective is to use cost information from the EU countries as proxy to estimate the cost of treating TB. The cost predictions provide essential information that is part of the evidence needed for budgeting and financing the health care facilities of TB services, especially with respect to avoiding under-estimation of the cost of TB-MDR treatment.

Keywords

1. Introduction

Tuberculosis (TB) has remerged as a threat to global health and is considered to be the second most common cause of death due to infectious diseases after HIV/AIDS (Human Immunodeficiency Virus/Acquired Immuno-

---

*Corresponding author.

deficiency Syndrome) globally. In 2012, approximately 8.6 million new TB cases and 1.3 million TB-related deaths were reported including 0.17 million death due to multi drug resistance TB [1]. It has been evidenced that majority of the TB related deaths are reported from low and middle income countries [1]. The largest number of cases was reported from Asia (58%), and Africa (27%). Smaller proportions were reported from the Eastern Mediterranean Region (8%), European Region (4%) and The Americas (3%). The five countries that rank first to fifth in terms of total numbers of incident cases in 2012 are India (1.6 - 2.4 million), China (1.0 - 1.6 million), South Africa (0.38 - 0.57 million), Indonesia (0.34 - 0.52 million) and Pakistan (0.42 - 0.60 million). India and China alone accounted for an estimated 38% of TB cases worldwide [1]. However, prevalence and mortality rates appear to be falling in all the six regions reported by [1].

Disturbingly, there were an estimated 450,000 cases of multidrug-resistant TB (MDR-TB) in 2012 with an estimated 170,000 deaths [1]. Overall 3.6% of newly diagnosed and 20% of the previously treated cases had MDRTB. Globally, 92 countries had reported cases of extensively drug-resistant tuberculosis (XDR-TB) by the end of 2012 [1]. These last figures are reason for considerable concern and highlight a potential threat to our ability to treat tuberculosis, both in individual patients and in the context of a treatment program [2].

According to the WHO the incidence of TB all forms, smear positive TB and deaths from TB each rose by 6.2% between 1990 and 2004, but the prevalence declined. These figures suggest that the notification rates of TB in Saudi Arabia increased during this period, as did the successful treatment rate, such that a decrease in the total number of cases was seen at the same time as the incidence and death rates increased [1] [2]. Despite the noticeable improvements on notification and successful treatment rates yet Saudi Arabia has a moderate infection rate in comparison with the other countries [3]. **Figure 1** shows the TB incidence from 1992 to 2010. Note the sharp decline in the incidence starting from 2000, which may be attributed to the control measures (regulations related to mandatory TB testing, treatment protocol, and hospitalization of TB MDR patients) imposed by the country's Ministry of Health (MOH).

In the Kingdom of Saudi Arabia, there are up to nine million migrants mainly from TB endemic regions, in South/South East Asia and Africa. In addition, annually 3 - 6 million pilgrims are visiting the holy cities located in the western region of the Kingdom for performing the holy rituals, Hajj and Umrah. Interestingly, the majority of these pilgrims are coming from TB endemic areas of Asia and Africa. There are differences in the rate of TB infection between different regions of the country. For instance in Jeddah (sea and air ports for pilgrims) the infection rate can reach up to 64 cases per 100,000 compared with 32 per 100,000 in Riyadh (Central). The higher rate in Jeddah may be attributed to the inflow of pilgrims.

The published data depict that TB in Saudi Arabia is still not fully controlled despite the huge efforts exerted by the MOH under the National TB Programs to eradicate the disease even with the implementation of Direct Observed Therapy (DOTS). However in the last 10 years, the mortality rates among TB patients in the country showed a decreasing trend (7.2% to 6.1%) among Saudis and a steady trend among non-Saudis [3]. A distribution of the number of TB cases by province over the years 2000-2009 in KSA was given in Table 6 of reference [3]. A graphical representation of the data is given **Figure 2**.

From the available data we would like to build a prediction model to provide projections for the future burden of the disease in KSA. This model may help public health officials determine the most appropriate intervention strategy to reduce the burden of diseases. The immediate benefits that can be gained from this predictive model are:

- With the extraneous available cost data, the model helps in evaluating the economic burden for TB and MDR treatments. We shall develop a reliable model, with high prediction accuracy to predict the TB counts in KSA.
- Many TB patients may develop Multi-Drug-Resistance (MDR) and this would substantially increase the cost of therapy. Predicting the MDR fraction of patients from the total TB count is required to provide a more credible estimate of the economic burden of the disease.

The most likely major cost saving for TB treatment is the reduction in MDR admissions [4], hence reducing the risk of hospital admission, and avoiding toxicity resulting from giving higher doses to patients.

This paper is structured as follows: In Section 2 we use the data published in [3] to build a time series fixed effect model from which we predict future counts of TB cases. In Section 3 we analyze the EU cost data published in [5] and establish a relationship between cost of treatment, the number of TB cases and the Gross Domestic Product (GDP) of each country, and in Section 4 we combine the results obtained from Sections 3 and 4 to predict the treatment cost of both TB and MDR cases. A general discussion will be taken up in Section 5.
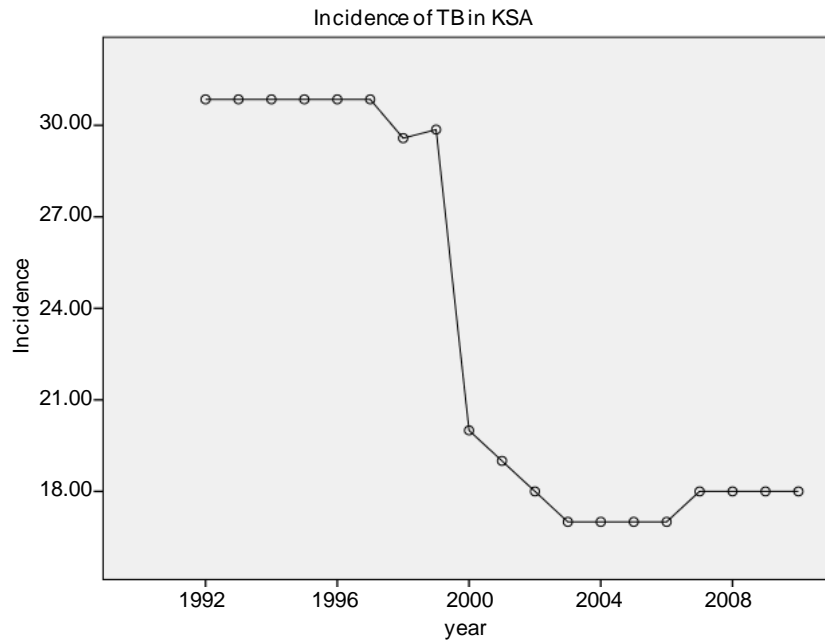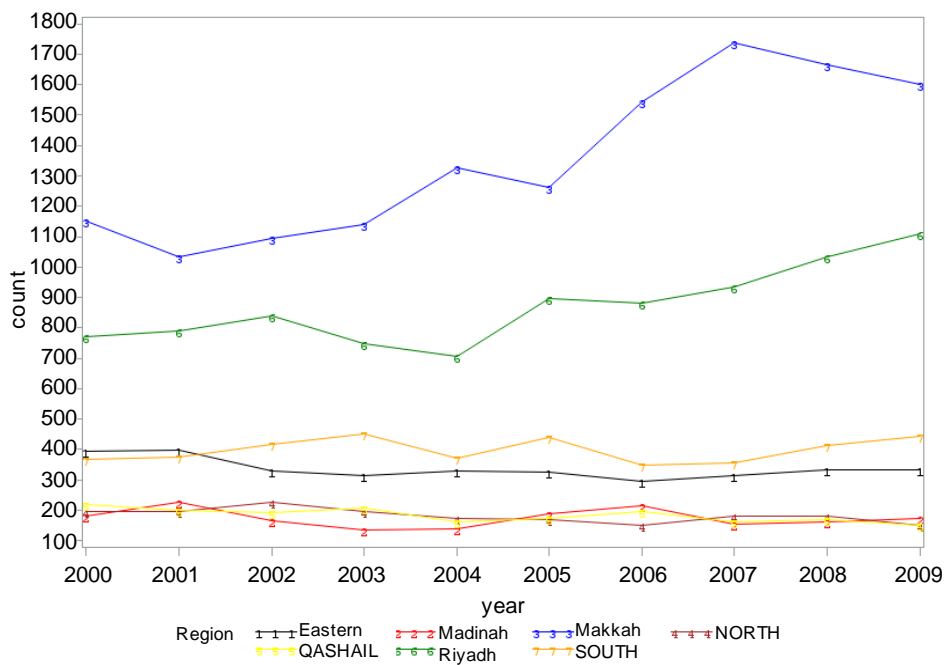
Figure 1. Incidence of TB in KSA.

Figure 2. TB count in Saudi Arabia (2000-2009) by province (REF).

2. TB Counts Prediction: Generalized Linear Regression Models

The data presented in **Figure 2**, prevalence data in each province given in [3] show the trend over time. From **Figure 2**, there seems to be a clustering of the data in three sub-groups. The first is Makkah, the second is the Riyadh Region and all other regions constituted as a third cluster. Regardless of this apparent sub-group clustering, the natural physical grouping of the data in the 7 regions will be kept in-tact, and regions will be explicitly included as 6 fixed covariates in the proposed modeling strategy. The available data are presented as time series of counts. Variability over time and region (province) must be accounted for in the process of building a predic-

tive model. Unlike most regression techniques, time series methods acknowledge the correlation structure of the data. For surveillance data which are generally collected monthly, or yearly the principal correlations are auto-correlations at a lag of one time period, or serial correlation [6]. When time series data are available over a relatively long period of time, it is important to estimate the trend and seasonal components as the auto-correlation structure can only be identified by using a stationary time series. Failure to account properly for the auto-correlation will result in a miss-specified model which may have bias in the estimated effects and prediction intervals which are too narrow.

A major attraction of longitudinal data is the ability to control for all stable covariates, without actually including them in a regression equation. In general, this is accomplished by using only within-regional variation to estimate the parameters, and then averaging the estimates over regions. Regression models for accomplishing this are often called fixed-effects models [7] [8].

Fixed-effects have been developed for a variety of different data types and models, including the Poisson regression models for count data [7]. The fixed-effects Poisson regression model for longitudinal data has been described in detail by Cameron *et al.* [9]. Let the dependent variable y_{it} denote the TB count at time t for region i $(i = 1, \cdots, n)$ and over time $(t = 1, \cdots, T)$. It is assumed that y_{it} to have a Poisson distribution with parameter μ_{it} which, in turn, depends on a vector of exogenous variables x_{it} according to the log-linear link function [9]:

$$\mu_{it} = \delta_i + \beta x_{it} \tag{1}$$

One way to estimate the parameter of this model is to do conventional Poisson regression by maximum likelihood, including dummy variables for all 7 regions (less one) to directly estimate the fixed effects.

For Poisson regression, two estimation methods—unconditional maximization of the likelihood and conditional likelihood—always yield identical estimates for β and the associated covariance matrix [9]. For the fixed-effects Poisson regression model we have the restriction that the mean of each count must equal its variance:

$$E(y_{it}) = \operatorname{var}(y_{it}) = \mu_{it} \tag{2}$$

The probability distribution of y_{it} is given by:

$$P(Y_{it} = y) = \exp(-\mu_{it}) \mu_{it}^y / y! \tag{3}$$

In many data sets, however there may be additional heterogeneity not accounted for by the model. A potential problem with our data is that there is still some evidence of over-dispersion in the data. Substantial departures from the mean variance ratio could indicate a problem with the model specification, and also suggest that the estimated standard errors may be downwardly biased.

One way to deal with the problem of over-dispersion is to alternatively assume that y_{it} has a negative binomial distribution (NBD), which can be regarded as a generalization of the Poisson distribution with an additional parameter (dispersion parameter) allowing the variance to exceed the mean [9]. The probability distribution of NBD is given by:

$$f(y_{it} \mid \mu_{it}, \theta) = \frac{\Gamma(\mu_{it} + y_{it})}{\Gamma(\mu_{it}) \Gamma(y_{it} + 1)} \left(\frac{\theta}{1+\theta}\right)^{y_{it}} \left(\frac{1}{1+\theta}\right)^{\mu_{it}} \tag{4}$$

The parameter θ is assumed to be constant over time for each region or province μ_{it} depends on covariates through the transformation:

$$\ln \mu_{it} = \beta x_{it} \tag{5}$$

The mean and variance of y_{it} are given by

$$\begin{aligned} E(y_{it}) &= \mu_{it}, \\ \operatorname{var}(y_{it}) &= (1 + \theta \mu_{it}) \mu_{it}. \end{aligned} \tag{6}$$

Therefore, when $\theta = 0$, the NB distribution reduces to the Poisson distribution. For the time series regional data graphed in **Figure 2**, we have two fixed effects, the year, and the province. Therefore the parameters vector β has three components. We use dummy variables for the fixed effects representing regions, and a continuous

covariate representing time, we fitted the count data with both models using the SAS software V9.3. Results of fitting the Poisson and NB regression models are summarized in **Table 1**. The goodness of fit measure is the scaled deviance. The closer the scaled deviance to 1, the better the fit is. The Akaike Information Criterion (AIC) may also be used as measure of goodness of fit as well (the smaller the AIC the better the fit is). The *p*-value is summary statement of the omnibus test that, the data follow the underlined model. A small *p*-value indicates that the Poisson model is not supported by the data. Clearly the NB is superior to the Poisson model.

The NB fitted equation, under the log-link is given by:

$$\log\left(\text{TB\_count}\right) = -2.439 + 0.0048\left(\text{year}\right) - 2.046\left(X_1\right) - 0.441\left(X_2\right) - 1.389\left(X_3\right) - 1.995\left(X_4\right)$$
$$- 2.002\left(X_5\right) - 1.228\left(X_6\right), \tag{7}$$

where

$$X_1 = \begin{bmatrix} 1 & \text{if Medinah} \\ 0 & \text{else} \end{bmatrix},$$

$$X_2 = \begin{bmatrix} 1 & \text{if Riyadh} \\ 0 & \text{else} \end{bmatrix},$$

$$X_3 = \begin{bmatrix} 1 & \text{if Eastern Province} \\ 0 & \text{else} \end{bmatrix},$$

$$X_4 = \begin{bmatrix} 1 & \text{if Qasim \& Hail} \\ 0 & \text{else} \end{bmatrix},$$

$$X_5 = \begin{bmatrix} 1 & \text{if Northern territory}^* \\ 0 & \text{else} \end{bmatrix},$$

$$X_6 = \begin{bmatrix} 1 & \text{if Soutthern territory}^* \\ 0 & \text{else} \end{bmatrix}.$$

*Northern territory includes North, Tabuk, and Al-Jouf. Southern territory includes Aseer, Jizan, Naran, and Al-Baha. The Makkah Region is the reference.

For example the predicted count in Makkah in the year 2020 is:

$$\exp\left[-2.439 + 0.0048\left(2020\right)\right] = 1418 \text{ TB cases}$$

All covariates are significant with *p*-value < 0.0001.

For a model to be used as predictive tool, the observed counts should be as close as possible to the model based predictive counts. This basically is the concept of goodness-of-fit, which also means that the model has to be internally valid. The Pearson's Chi-square is often used for this purpose. However, this test is affected by one or two discrepancies between observed and expected counts, even if the rest of the expected counts are quite close to their corresponding observed counts. We alternatively use the Concordance Correlation Coefficient (CCC) ρ_c [10] to assess the goodness of fit of the NBM, or its internal validity. The CCC is given by:

$$\rho_c = \frac{2r\text{SD}\left(o\right)\text{SD}\left(p\right)}{\text{variance}\left(o\right) + \text{variance}\left(p\right) + \left(\text{mean}\left(o\right) - \text{mean}\left(p\right)\right)^2} \tag{8}$$

Table 1. Fitting Poisson and NB regression models.

Output	Poisson	Negative binomial
Deviance	557.75	62.05
DF	62	62
Scaled deviance	8.996	1.114
AIC	1115.569	761.44
p-value	0.00001×10^{-4}	0.474

SD(o) and SD(p) are respectively the standard deviations of the observed and the NB predicted counts, and r is the Pearson's correlation between the two sets of counts. The observed and predicted totals are depicted in **Table 2**.

Direct computation from the data in **Table 3** gives $\rho_c = 0.91$. This high value of the CCC confirms the high reproducibility of the model as a predicting tool.

Based on the NBM, the projections from the year 2011 to 2020 are given in the following **Table 3**.

3. Statistical Analysis of Cost of Treating TB in the EU

The main objective of this section is to review and summarize the available evidence on the global cost and the cost components of TB and MDR in the EU countries [5]. Comparing cost of illness studies is difficult because of the different definitions and methods used to measure and quantify cost, whether direct or indirect. The scope of the indirect cost measurement varies considerably across studies: some only include economically active individuals, but others include children and the elderly; most measure the time spent seeking treatment by the patient and caregiver and their loss of productive labor time due to illness.

Diel *et al*. [5] performed a systematic literature review, followed by the extraction of direct as well as indirect costs of TB. Direct costs include costs for the medical treatment of TB (medication, laboratory, hospitalization and outpatient visits) and indirect costs represent productivity loss because of TB-induced sick days off-work. They gathered information on several cost components based on data available in the year 2012:

1) Duration of hospitalization and inpatient cost per day.
2) Outpatient cost.
3) Cost of medication.
4) Cost due to loss of productivity.
5) Year in which the costs were assessed or collected.

The authors of [5] recognized the limitation of the review. Nevertheless, despite the limitations, the best available cost data are part of the evidence needed for budgeting and financing the expansion of TB services, especially with respect to scaling-up MDR-TB treatment. In their review, the authors emphasize the fact that lack of sufficient information from many countries makes extrapolation a necessity.

It should be noted that the recent WHO global tuberculosis report [11] data may be utilized to identify the major factors that affect the total cost levels. These are the total TB count and cost of treatment, percentage of MDR and its treatment cost. In reviewing the literature on cost of treating TB, we discovered that there is a wide variation in allocation of treatment cost that is attributed mainly to the annual household income. To strengthen the analysis we augmented the data with information regarding the Gross Domestic Product [12] of each country in the data base.

Our proposed cost estimation proceeded in three steps:

1) In the first step, we split the EU countries into two groups. The first group included the founding nations of the EU, and this includes 16 countries (Austria, Belgium, Cyprus, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxemburg, Malta, Portugal, Spain, Sweden, and UK). The second group includes countries that joined the union at later dates, and this group included 10 countries (Bulgaria, Czeck R. Estonia, Hungary, Latvia, Lithuania, Poland, Romania, Slovakia, Slovenia), most of which were the former Eastern-Bloc Nations. To evaluate the accuracy of the grouping, we used the ROC to classify the EU nations into two categories using as a predictor. The ROC curve is given in **Figure 3**.

Table 2. Goodness of fit of the NBM when fitted to the current data.

Year + 2000	0	1	2	3	4	5	6	7	8	9
Observed count (o)	3284	3221	3266	3194	3207	3460	3637	3847	3957	3964
Predicted count (p)	3098	3182	3268	3356	3447	3540	3636	3734	3835	3939

Table 3. Projected future total of TB counts in KSA based on the fixed-effects NBM.

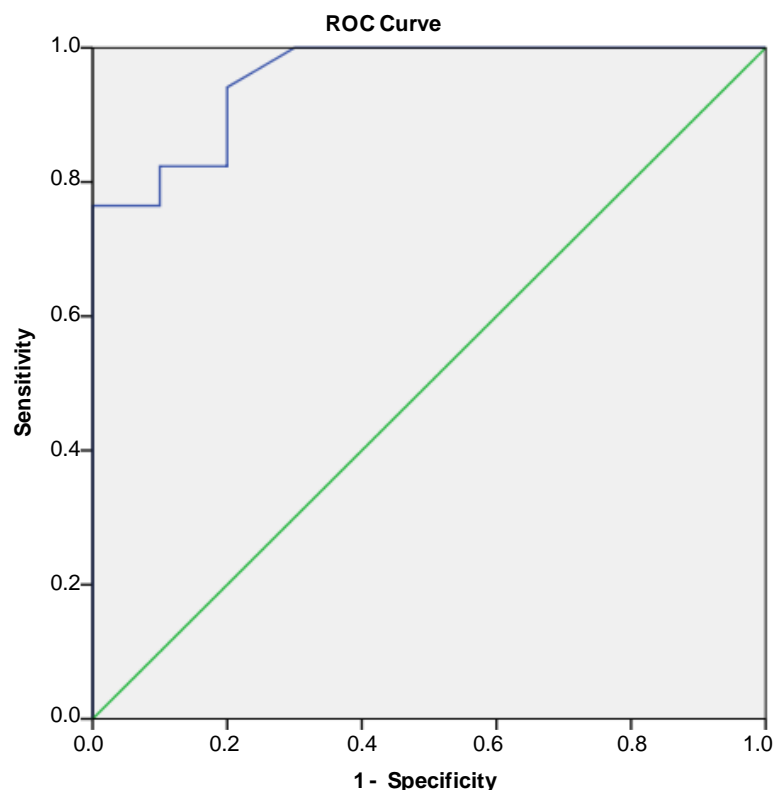| Year + 2000	11	12	13	14	15	16	17	18	19	20
Projected count	4171	4286	4403	4523	4647	4774	4905	5039	5177	5319

Figure 3. The receiver operating characteristic (ROC) curve to classify each of the EU country according to its GDP.

The area under the AUC is 0.956 ± 0.036 (*p*-value < 0.00001), and the cut-off point is US \$24,800. At this point the sensitivity and specificity of our classification scheme were given respectively as 76% and 80%. The coefficient of agreement "kappa" between the observed and the predicted classifications was 78%, which according to the benchmarks set by Landis and Koch [13] is substantial.

2) In **Table 4**, we produced some relevant summary statistics for the two groups. The two groups differ significantly with respect to the GDP/per capita and the log-cost of treatment per TB case. Thereafter we looked at the relationship between the total cost of treating TB and the number of TB cases. It is evident from **Figure 4** evident, which is the scatter plot of the cost of treatment and the number of TB cases, that there is a quadratic relationship between the total cost and the number of TB cases (**Figure 4**). When we fitted a quadratic regression model to the data, the relationship is given Equation (8).
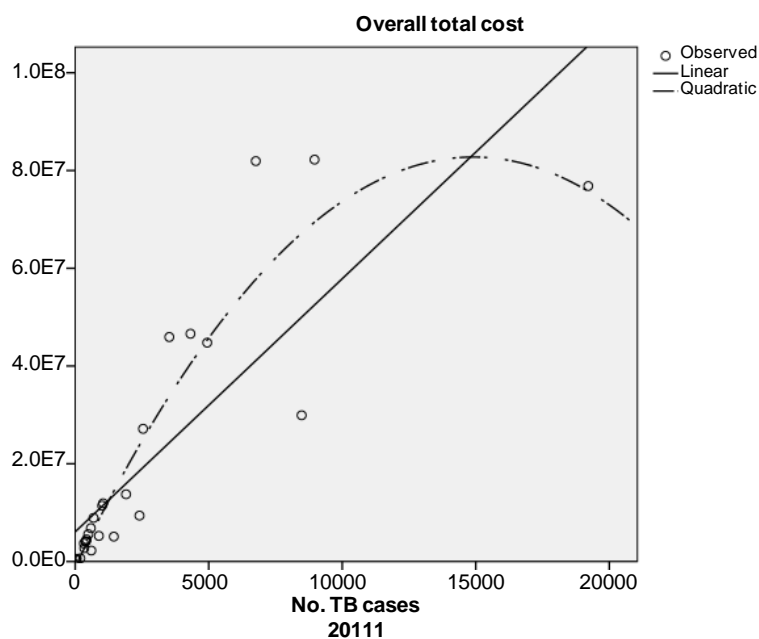
$$\text{Total cost}(2011) = -1075745.5 + 11248.9\ (\text{TB}) - 0.377\ (\text{TB}^2) \tag{8}$$

In (8) TB is the total count in 2011 The percent of explained variation in total cost that attributed to the total number of TB cases is measured by $R^2 = 0.846$, indicating an excellent fit.

In an interesting review [14] it was reported that there is a significant relationship between the treatment cost per TB case and the annual household income, a variable that is not measured in the data base that we are analyzing from [5]. However, we may use the GDP as a proxy variable [15] to predict the cost of treatment. Proxy variables are occasionally used in econometric analysis when important independent predictors are not available. We attempted to model the relationship between cost of treatment per TB case and the GDP as a proxy variable, but the relationship was not statistically significant. However, when we dichotomized the GDP at the cut-off vale \$24,900, as mentioned earlier, we created two classes of countries, thereafter. Using this categorical variable as a predictor, we were able to establish a statistically significant relationship between the country-class and the cost per TB case on the log scale. This approach may be controversial, but we claim that this is a case of zero-causal effect but positive predictive comparison between the two classes of countries [15]-[17]. This regression equation is given in (9)

Table 4. Comparison between the two groups of EU countries included in the study.

	1) $n = (16)$	2) $n = (10)$	*p*-value
Number of TB	2124 ± 2679	3568 ± 6012	0.39
Cases 2011 Log (cost-per TB case)	9.32 ± 0.113	8.507 ± 0.408	<0.0001
GDP/per capita in US $	37,264 ± 12,589	20,430 ± 5206	<0.0001
% MDR among TB	0.02 ± 0.02	0.06 ± 0.08	<0.00002

Figure 4. Scatter plot of the total cost of treatment of TB and the number of cases.

$$\text{Log}\left(\text{cost-per TB case}\right) = 10.134 - 0.813\left(\text{country class}\right), \tag{9}$$

where country class = 1, if a country belongs to the first 16 nations, and zero otherwise. The negative value of the regression coefficient means that the group of countries in class = 1 have on the average lower cost of treatment per TB case, relative to the other 10 countries in the sample.

4. Cost Prediction in KSA

Since the EU cost data are for the year 2011, the projected number of TB cases in 2011 in KSA, based on the NBM is 4171. It has been reported that the GDP in KSA is 29,600 US. This means that KSA belongs to country-class = 1.

Therefore the estimated total cost of TB in KSA in 2011

$$= 4171 \times \exp\left(10.134 - 0.813\right) = 46,590,681 \text{ US or } 174,715,053 \text{ SR}.$$

This is a baseline cost, and it in be adjusted upward to offset the effect of inflation, or downward if a cheaper drugs are available.

On the other hand, from the EU data [5] the relationship between the total cost of treatment of MDR-TB is linearly related to the number of MDR-TB cases (scatter plot is **Figure 5**). The regression equation is:

$$\text{Total cost due to MDR} = 1852636.963 + 25829.855\left(\text{No. MDR} - \text{TB}\right) \tag{10}$$

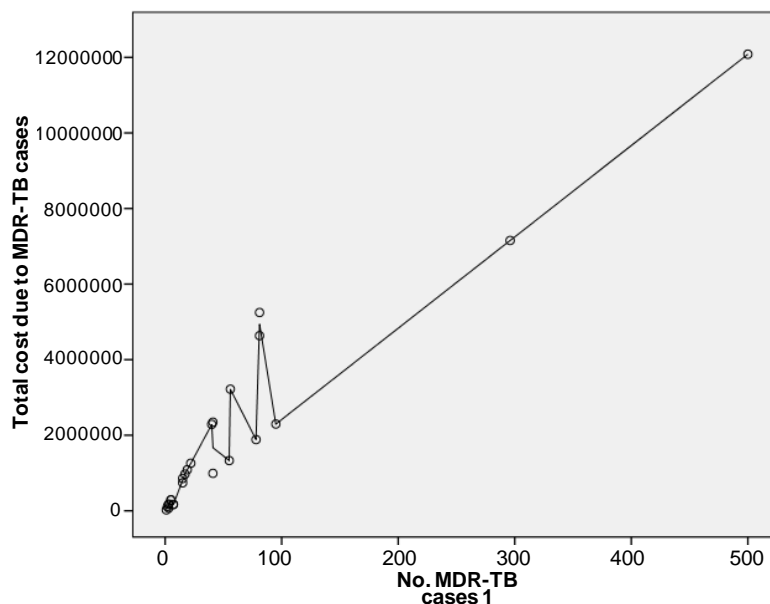The fitted equation is excellent since the $R^2 = 0.924$.

Figure 5. Linear relationship between the total costs of treatment MDR-TB and the total number of MDR-TB cases.

Now, if the percentage of MDR in KSA in 2011 was 0.045, this means that the projected number of MDR-TB cases in KSA is $0.045 \times 4171 = 188$ MDR cases.

The projected cost due to MDR in KSA is therefore expected to be:

$$1852636.963 + 25829.855(188) = 6,7086,50 \text{ US or } 25,157,436 \text{ SR}.$$

This means, on average MDR cost $133,816 \text{ SR} (25,157,436/188)$ per case, while a TB-case costs $174,715,053/4171 = 41,888 \text{ SR}$.

The above calculations show that although MDR-TB is a small fraction of the total TB cases, it consumes a huge proportion of the total budget allocated for the treatment of TB. It is therefore aids the policy makers in their disease management strategy.

5. Discussion

Despite declining global TB incidence, this infectious disease remains a major public health issue, particularly in KSA. What exacerbates the problem is the large number of migrant workers who come from countries with high TB burden. In several EU countries where there was a low incidence of TB, the rate of decline had slowed down [18]. Drug resistance and HIV co-infection are the main factors affecting the trend [19]. From the regression analysis of the EU cost data, we found that the total cost of treating TB was a quadratic function of the TB count. Therefore a reliable estimate of TB count is needed in order to accurately estimate the total treatment cost, moreover the reliability of the prediction would definitely improve when we use longer time series data. The NBM was proved to be a good predicting tool with and its goodness of fit to the data was excellent. Because we lacked information about the treatment cost in KSA, we relied on the published EU cost data and the GDP of the EU countries to establish a relationship between the cost-per TB case and the GDP-dependent class category of the country, and from this relationship we produced a base-line estimate of the total cost of TB treatment in KSA. This algorithm has two limitations. Firstly, the cost structure, as reported in the meta-analysis paper [5], varies from one country to the other, even within the EU countries, and naturally this variation should encompass KSA. The second is related to the use of proxy information in the prediction process. This approach has theoretical challenges that could not be discussed in this paper.

Acknowledgements

References

[1]   (2014) WHO TB Report 2013. http://www.who.int/tb/publications/global_report/en/

[2]   Donald, P.R. and Van Helden, P.D. (2009) Global Burden of TB-Combating Drug Resistance in Difficult Times. *NEJM*, **360**, 23. http://dx.doi.org/10.1056/NEJMp0903806

[3]   Abouzaid, M.S., Zumla, A.I., Felemban, S., Alotaibi, B., O'Grady, J. and Memish, Z.A. (2012) Tuberculosis Trends in Saudi and Non-Saudi in the Kingdom of Saudi Arabia—10 Year Retrospective Study (2000-2009). *PLoS ONE*, **7**, e39478. http://dx.doi.org/10.1371/journal.pone.0039478

[4]   Al-Hajoj, S.A. (2010) Tuberculosis in Saudi Arabia: Can We Change the Way We Deal with the Disease? *Journal of Infection and Public Health*, **3**, 17-24. http://dx.doi.org/10.1016/j.jiph.2009.12.001

[5]   Diel, R., Vandeputte, J., de Vries, G., Stillo, J., Wanlim, M. and Nienhaus, A. (2014) Costs of Tuberculosis Disease in the EU—A Systematic Analysis and Cost Calculation. *European Respiratory Journal*, **43**, 554-565. http://dx.doi.org/10.1183/09031936.00079413

[6]   Fitzmaurice, G.M., Laird, N.M. and Ware, J. (2011) Applied Longitudinal Analysis. 2nd Edition, Wiley, New York.

[7]   Palmgren, J. (1981) The Fisher Information Matrix for Log-Linear Models Arguing Conditionally in the Observed Explanatory Variables. *Biometrika*, **68**, 563-566.

[8]   Allison, P.D. (1996) Fixed Effects Partial Likelihood for Repeated Events. *Sociological Methods & Research*, **25**, 207-222. http://dx.doi.org/10.1177/0049124196025002003

[9]   Cameron, A.C. and Trivedi, P.K. (1998) Regression Analysis of Count Data. Cambridge University Press, Cambridge. http://dx.doi.org/10.1017/CBO9780511814365

[10]   Lin, L.I. (1989) A Concordance Correlation Coefficient to Evaluate Reproducibility. *Biometric*, **45**, 255-268. http://dx.doi.org/10.2307/2532051

[11]   World Health Organization (2012) Global Tuberculosis Report 2012. WHO, Geneva.

[12]   The CIA World Fact-Book (2014) https://www.cia.gov/library/publications

[13]   Landis, J.R. and Koch, G.G. (1977) The Measurement of Observer Agreement for Categorical Data. *Biometrics*, **33**, 159-174. http://dx.doi.org/10.2307/2529310

[14]   Russell, S. (2004) The Economic Burden for Household in Developing Countries: A Review of Studies Focusing on Malaria, Tuberculosis, and Human Immunodeficiency Virus/Acquired Immunodeficiency Syndrome. *American Journal of Tropical Medicine and Hygiene*, **71**, 147-155.

[15]   Stahlecker, P. and Trenkler, G. (1993) Some Further Results on the Use of Proxy Variables in Prediction. *The Review of Economics Statistics*, **75**, 707-711. http://dx.doi.org/10.2307/2110026

[16]   Sobel, M.E. (2006) What Do Randomized Studies of Housing Mobility Demonstrate? Causal Inference in the Face of Interference. *Journal of the American Statistical Association*, **101**, 1398-1407. http://dx.doi.org/10.1198/016214506000000636

[17]   Sobel, M.E. (2008) Identification of Causal Parameters in Randomized Studies with Mediating Variables. *Journal of Educational and Behavioral Statistics*, **33**, 230-251.

[18]   Wright, A., Zignol, M., Van Deun, A., Falzon, D., Gerdes, S.R., Feldman, K., *et al*. (2009) Epidemiology of Antituberculosis Drug Resistance 2002-07: An Updated Analysis of the Global Project on Anti-Tuberculosis Drug Resistance Surveillance. *The Lancet*, **373**, 1861-1873.

[19]   Broekman, J.F., Migliori, G.B., Rieder, H.L., Lees, J., Ruutu, P., Loddenkemper, R., *et al*. (2002) European Framework for Tuberculosis Control and Elimination in Countries with Low Incidence. *European Respiratory Journal*, **19**, 765-775.

Scientific Research Publishing (SCIRP) is one of the largest Open Access journal publishers. It is currently publishing more than 200 open access, online, peer-reviewed journals covering a wide range of academic disciplines. SCIRP serves the worldwide academic communities and contributes to the progress and application of science with its publication.

Other selected journals from SCIRP are listed as below. Submit your manuscript to us via either submit@scirp.org or Online Submission Portal.