Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v17n2p382

Relationships between Smoking Rate and COVID-19 Death Rate when Both Variables are Subject to Errors: Arab World Case Study

By AL-Ananzeh, and Al-Nasser

15 October 2024

Electronic Journal of Applied Statistical Analysis
Vol. 17, Issue 02, October 2024, 382-391
DOI: 10.1285/i20705948v17n2p382

Relationships between Smoking Rate and COVID-19 Death Rate when Both Variables are Subject to Errors: Arab World Case Study

Abeer K. AL-Ananzeh[a] and Amjad D. Al-Nasser*[a]

[a]*Department of Statistics, Faculty of Science, Yarmouk University, Jordan.,*

15 October 2024

While there is a growing body of research on the relationship between smoking and COVID-19 outcomes, there is a lack of comprehensive studies focusing specifically on the Arab world. Understanding the potential impact of smoking on COVID-19 death rates in this region is crucial for informing public health interventions and policies. This article discusses the relationship between the smoking rate and the COVID-19 death rate in the Arab world. The data was gathered from the World Health Organization (WHO) records. The measurement error model (MEM) is used to study the relationship between smoking rate and COVID-19, considering the errors in both variables. Different estimation methods including the maximum likelihood estimation (MLE), Wald-type methods, and repeated median estimation methods, were used to fit the MEM model to the data. The results showed that Jordan and Lebanon, respectively, had the highest smoking rates while Yemen, Egypt, and Algeria, respectively, had the highest COVID-19 death within the Arab countries. Fitting the data to MEM, the data analysis showed that Smoking has a positive effect on COVID-19.

keywords: Measurement Error Model, Grouping Method, Repeated Median, COVID-19, Smoking Rate.

*Corresponding authors: amjadyu@yahoo.com

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

1 Introduction

Without a shadow of a doubt, a positive correlation was found between tobacco use and the risk of COVID-19 disease. Cigarette smoking is one of the most preventable causes of morbidity and mortality in humans (Maziak et al., 2014). Smoking increases the risks of many serious health conditions, including cardiovascular disease, cancers of the lung, oropharynx, larynx, and pancreas, low birth weight, and chronic obstructive pulmonary disease (Alzoubi et al., 2010). Smoking causes eight million deaths worldwide per year. More than 80% of all smokers now live in countries with low or middle incomes. More than 1.3 billion people smoke tobacco, which is nearly 22.3% of the world's population in 2022 (Organization, 2023). On the other hand, the coronavirus disease 2019 (COVID-19) pandemic, an infection caused by the severe acute respiratory syndrome corona virus (SARS-CoV-2), has led to more than 7,010,681 deaths worldwide (WHO, 2023). COVID-19, as well as smoking; has detrimental effects on the respiratory system, causing harm to the airways and hindering lung function (Maziak et al., 2014). The evidence suggests that smokers are more vulnerable to lung infection, and COVID-19 is no exception (Guo, 2020).

Both variables are considered common risk factors for lung functions, many respiratory infections including SARS, MERS-Cov; and immune systems. Many researchers investigated the relationship between both variables in their research. Gaiha et al. (2020), showed that COVID-19 has a significant association with smoking cigarettes or e-cigarettes. Also, Kashyap et al. (2020), showed that active smoking is associated with increased severity of disease and death in hospitalized COVID-19 patients. Usman et al. (2021), introduced a good review of the relationship between smoking and COVID-19 and highlighted current knowledge gaps. Although certain studies suggest a correlation between smoking and heightened vulnerability to COVID-19, others show a potential protective effect among smokers. He et al. (2022), suggested that many studies should be conducted to clarify the relationship between smoking and COVID-19 since the relationship between these variables is still unclear. Also, Lippi and Henry (2020), based on a meta-analysis that consists of five different research studies, claimed that active smoking does not seem to be significantly associated with an enhanced risk of progressing toward severe disease in COVID-19. In a cross-sectional study, Hoballah et al. (2022), showed sex differences in the association between tobacco smoking and COVID-19 severity. They claimed that tobacco smoking was not associated with SARS-CoV-2 infection severity among female patients, however, tobacco smoking, particularly water pipe, was found to be associated with infection severity among males. These conflicting results have sparked a continuous debate within the scientific community about the extent of smokers' susceptibility to COVID-19 infections. In this article, we are interested in studying the relationships between smoking rate and COVID-19 in the Arab world, considering the measurement error in both variables.

The recent of this article is organized as follows: in Section 2, the research model is given and discussed; while Section 3 discusses some well-known estimation methods used to fit the MEM. Data analysis and fitting the relationships between the smoking death rate and the death rate caused by COVID-19 using different estimation methods are given in Section 4. This article ends with a concluding remark in Section 5.

2 The Model

In this study, the measurement error will be considered in both variables of interest. Therefore, to study the relationships between two variables or more the the measurement error model (MEM) will be considered (Kendall and Stuart, 1979; Fuller, 1987; Carroll et al., 2006). MEM has been successfully applied to several research areas such as, but not limited to: engineering, econometrics, Quality, environment, ecology, and health care. A complete description of the MEM model can be found in many research articles (Cheng and Van Ness, 1999; Al-Nasser et al., 2005; Al-Radaideh et al., 2010; Eidous and Al-Nasser, 2018; Al-Nasser et al., 2016; Al Dibi'i and Al-Nasser, 2019).

The mathematical MEM can be formulated based on a bivariate random sample of size n $\{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, using two mathematical equations:

$$y_i = \beta_0 + \beta_1 \xi_i + \varepsilon_i \quad \text{for } i = 1, 2, \ldots, n$$

$$x_i = \xi_i + \delta_i \quad \text{for } i = 1, 2, \ldots, n$$

where ξ is the latent (unobservable) variable that cannot be measured directly, β_0 and β_1 are unknown parameters. The error terms δ and ε are assumed to be independently distributed (Al-Nasser, 2011; Cheng and Van Ness, 1999; Kendall and Stuart, 1979). The main problem in MEM concerns model identifiability. There are six different assumptions (any of them) that can be used to make the model identifiable (Al-Nasser, 2005):

1. The ratio of error variance; $\lambda = \frac{\sigma_\varepsilon^2}{\sigma_\delta^2}$ is known.
2. The reliability ratio; $K_\xi = \frac{\sigma_\xi^2}{\sigma_\delta^2 + \sigma_\xi^2}$ is known.
3. σ_ε^2 is known.
4. σ_δ^2 is known.
5. σ_δ^2 and σ_ε^2 are known.
6. The intercept β_0 is known and $E(X) \neq 0$.

3 Estimation Methods to MEM

Several estimation methods have been suggested to estimate the unknown parameters of the MEM using the above six side assumptions, such as the maximum likelihood estimation method, the repeated median method, and the Grouping method.

3.1 Maximum Likelihood Estimation Method

The maximum likelihood estimation (MLE) method leads to different solutions based on the presumptions (Fuller, 1987; Cheng and Van Ness, 1999). Table 1 gives the different assumptions under the structural MEM for each assumption.

Table 1: MEM Intercept and Slope Parameter Estimators.

Known Assumption	$\hat{\beta}_0$	$\hat{\beta}_1$
λ	$\hat{\beta}_{0,1}$	$\hat{\beta}_{1,1}$
K_ξ	$\hat{\beta}_{0,2}$	$\hat{\beta}_{1,2}$
σ_δ^2	$\hat{\beta}_{0,3}$	$\hat{\beta}_{1,3}$
σ_ε^2	$\hat{\beta}_{0,4}$	$\hat{\beta}_{1,4}$
σ_δ^2 and σ_ε^2	$\hat{\beta}_{0,5}$	$\hat{\beta}_{1,5}$
β_0	$\hat{\beta}_{0,6}$	$\hat{\beta}_{1,6}$

where

1. $\hat{\beta}_{0,i} = \bar{y}_n - \hat{\beta}_{1,i}\bar{x}_n; \quad i = 1, 2, 3, 4.$
2. $\hat{\beta}_{1,1} = \frac{(s_{yy(n)} - \lambda s_{xx(n)}) + \sqrt{(s_{yy(n)} - \lambda s_{xx(n)})^2 + 4\lambda s_{xy(n)}^2}}{2s_{xy(n)}}.$
3. $\hat{\beta}_{1,2} = \frac{s_{xy(n)}}{s_{xx(n)} K_\xi} = K_\xi^{-1} \frac{s_{xy(n)}}{s_{xx(n)}}.$
4. $\hat{\beta}_{1,3} = \frac{s_{xy(n)}}{s_{xx(n)} - \hat{\sigma}_\delta^2}$ assuming that $s_{xx(n)} > \sigma_\delta^2$.
5. $\hat{\beta}_{1,4} = \frac{(s_{yy(n)} - \hat{\sigma}_\varepsilon^2)}{s_{xy(n)}}$ assuming that $s_{yy(n)} > \hat{\sigma}_\varepsilon^2$.
6. $\hat{\beta}_{1,6} = \frac{\bar{y} - \beta_0}{\hat{\mu}} = \frac{\bar{y} - \beta_0}{\bar{x}_n}$; $\hat{\mu} = \bar{x}_n$ assuming that $\bar{x}_n \neq 0$.

3.2 Wald-Type Methods

The Wald-type estimation method has no assumption regarding the structure of error proposed by Wald (1940) and is called the grouping method. The procedure of this method consists of rearranging the observed x values in ascending order, then taking the related y values, and dividing the observations into equally sized subgroups (G1 and G2) and finding the mean for each group; the procedure can be described as follows (Al-Nasser, 2005):

1. Arrange the observations in ascending order based on the values of x_i.
2. Divide the data into two equal subgroups (the odd number of observations can be treated by omitting the median).

3. Compute the arithmetic means $(\bar{x}_1, \bar{y}_1)$ for the lower subgroups and $(\bar{x}_2, \bar{y}_2)$ for the upper subgroups.

4. Then the estimated values of the parameters are defined as:

$$\hat{\beta} = \frac{\bar{y}_2 - \bar{y}_1}{\bar{x}_2 - \bar{x}_1},$$

and

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}.$$

3.3 Repeated Median Estimation Method

The Thiel-type estimation method, proposed by Thiel (1950), begins by sorting either on the x variable or the y variable in ascending order. Then, the Thiel-type estimator is obtained via the following steps (Al-Nasser and Ebrahem, 2005):

1. **Step 1:** The data are ordered either by the x variable or the y variable.

2. **Step 2:** Find all possible pairs of observations if all x's are distinct.

3. **Step 3:** Calculate

$$B_{ij} = \frac{y_{[j]} - y_{[i]}}{x_{(j)} - x_{(i)}} \quad ; \quad i = 1, 2, \ldots, j-1, \quad j = 2, 3, \ldots, n$$

4. **Step 4:** The Thiel-type estimator is given by

$$\hat{\beta}_T = \text{median}(B_{ij})$$

4 Data

There are two variables considered in this article including the death rate because of the COVID-19 pandemic and the smoking rate; in the Arab countries. Table 2 shows the available data based on Organization (2023) reports. Formally WHO started tracking the coronavirus on January 29, 2020. However, they stopped tracking or reporting the virus as of April 13, 2024. Within this period over 704,753,890 cases and close to 7,010,681 deaths were reported globally, which is considered the last valid statistics. However, only the formal data reported about Arab countries by Organization (2021) is considered in this research article which consists of 14,273,889 cases and 174,045 deaths.

Table 2: Rate of Death COVID-19 and rate of smokers in the Arab states. Data as of 13 April 2024.

Country	Rate of Smokers (RS)	COVID-19 Death Rate (RC)
Algeria	21.20%	2.53%
Bahrain	15.00%	0.22%
Comoros	17.20%	1.77%
Egypt	24.70%	4.77%
Iraq	19.20%	1.03%
Jordan	35.60%	0.81%
Kuwait	19.90%	0.39%
Lebanon	34.30%	0.88%
Mauritania	9.50%	1.56%
Morocco	13.00%	1.27%
Oman	8.40%	1.16%
Qatar	12.50%	0.13%
Saudi Arabia	14.90%	1.15%
Tunisia	20.50%	2.55%
United Arab Emirates	9.00%	0.22%
Yemen	21.40%	18.07%

4.1 Descriptive Statistics

The simple descriptive statistic for the data is given in Table 2. Moreover, Figure 3 shows the box plot for both variables, from the box plot it could be noted that the number of confirmed cases with RC has two outliers (Egypt and Qatar). While the distribution of RS in the Arab world is approximately symmetric.

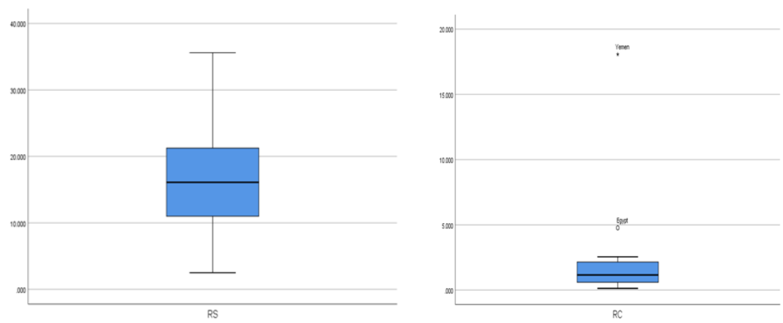

Figure 1: Box Plot for both variables.

For a better data description and classification both variables are presented by using the scatter plot (Figure 2), The Arab states are located into four quadrants based on the mean values of RS and RC.

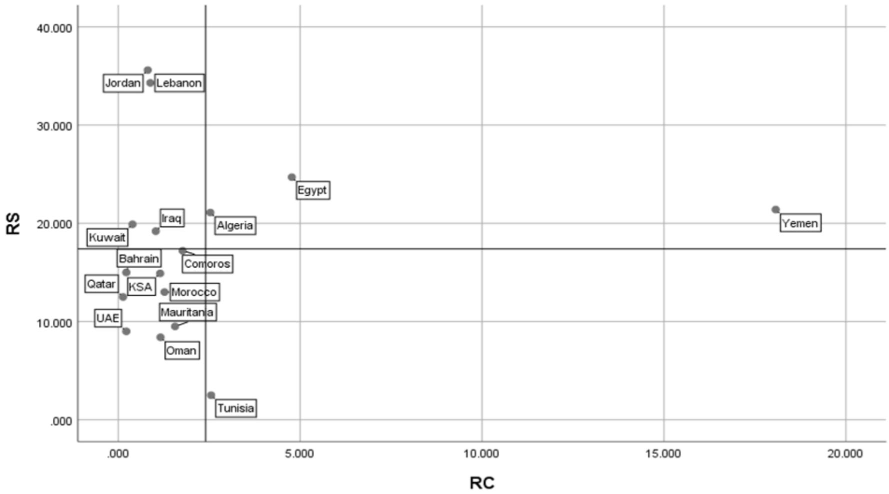

Figure 2: Scatter plot for RC and RS.

The classification results show that there are three Arab countries, including Algeria, Egypt, and Yemen are highly affected by both variables; however, Yemen is extremely affected by COVID-19. On the other hand, there are two Arab states (Jordan, and Lebanon) that are extremely affected by the smoking rate. Moreover, there are several Arab states not highly affected by both variables (Mauritania, Oman, Qatar, Bahrain, UAE, KSA, Morocco, and Comoros).

4.2 Fitting the relationship between RS and RC

Using the data, since there are outliers, it seems there is no linear relationship between RC (Y) and RS (X). Figure 3. MEM may help in fitting such data under the fact that there are errors in both variables. Hence the suggested MEM is:

$$RC = \beta_0 + \beta_1(\xi) + \varepsilon$$

$$RS = \xi + \delta$$

This can be rewritten in one general equation as:

$$RC = \beta_0 + \beta_1(RS - \delta) + \varepsilon$$

where α and β are the intercept and the slope, respectively, while δ and ε are the error terms. Fitting data to the MEM is given in Table 3 under different estimation methods. All methods showed that there is a positive relationship between both variables.

Table 3: MEM results using different methods

Method	**Known Assumption**	$\hat{\beta}_0$	$\hat{\beta}_1$
MLE	λ	0.9632	0.0830
	K_ξ	0.1826	0.1279
	σ_δ^2	-57.8395	3.4649
	σ_ε^2	1.2797	0.0648
	β_0 (known)	1	0.0809
Repeated Median		0.8612	0.0250
Wald-Type		-1.1139	0.2025

5 Concluding Remarks

This article seeks to address the gap in the literature by investigating the relationship between smoking rates and COVID-19 death rates in the Arab world, assuming that both variables are measured with errors. The data analyses in this article used a combination of descriptive statistics and linear relationship methods to examine smoking rates and COVID-19 death rates in the Arab world. Data on smoking prevalence obtained from national surveys and reports, while COVID-19 death rates will be sourced from official health (WHO) records and reports. The findings of this research showed that there is a positive relationship between smoking and the COVID-19 death rate. This research result is a shred of good evidence that can inform effective strategies for addressing the impact of smoking on COVID-19 outcomes in the region.

References

Al Dibi'i, R. and Al-Nasser, A. (2019). Fitting structural measurement error model using repetitive wald-type procedure. In *The 6th International Arab Conference on Mathematics and Computations (IACMC 2019)*, pages 24–26, Al Zarqa University, Jordan.

Al-Nasser, A. (2005). Entropy type estimator to simple linear measurement error models. *Austrian Journal of Statistics*, 34(3):283–294.

Al-Nasser, A. (2011). An information-theoretic alternative to maximum likelihood estimation method in ultra structural measurement error model. *Haccettape Journal of Mathematics and Statistics*, 40(3):469–481.

Al-Nasser, A., Al-Sliti, and Edous, O. (2016). New iterative am estimation procedure for fitting the simple linear measurement error models. *Electronic Journal of Applied Statistical Analysis*, 9(3):491–501.

Al-Nasser, A. and Ebrahem, M. (2005). A new nonparametric method for estimating the slope of simple linear measurement error model in the presence of outliers. *Pakistan Journal of Statistics*, 21(3):265–274.

Al-Nasser, A., Ebrahem, M., and Al-Masri, A. (2005). On the use of l-statistic in wald-type estimators for fitting a straight line when both variables are subject to error. *Pakistan Journal of Statistics*, 21(2).

Al-Radaideh, A., Al-Nasser, A., and Ciavolino, E. (2010). Efficient wald type estimators for simple linear measurement error model. *Asian Journal of Mathematics and Statistics*, 3(2):82–92.

Alzoubi, K., Azab, M., Khabour, O., Al-shamaila, A., Ayoub, N., Al-Omiri, M., Al-Nasser, A., Mhaidat, N., and Al-Azzam, S. (2010). Smoking-cessation practice guidelines: Awareness and implementation among medical teams. *International Journal of Pharmacy Practice*, 18(2):93–99.

Carroll, R., Ruppert, D., Stefanski, L., and Crainiceanu, C. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*. Chapman and Hall/CRC Press, London.

Cheng, C. L. and Van Ness, J. W. (1999). *Statistical Regression with Measurement Error*. Arnold, London.

Eidous, M. and Al-Nasser, A. (2018). On fitting measurement error model using sample entropy moments. *Statistica and Applicazioni*, XVI(1):19–29.

Fuller, W. A. (1987). *Measurement Error Models*. Wiley, New York.

Gaiha, S., Cheng, J., and Halpern-Felsher, B. (2020). Association between youth smoking, electronic cigarette use, and covid-19. *Journal of Adolescent Health*, 67(4):519–523.

Guo, F. (2020). Active smoking is associated with severity of coronavirus disease 2019 (covid-19): An update of a meta-analysis. *Tobacco Induced Diseases*, 18:37.

He, Y., He, Y., Hu, Q., Yang, S., Li, J., Liu, Y., and Hu, J. (2022). Association between smoking and covid-19 severity: A multicentre retrospective observational study. *Medicine*, 101(29):e29438.

Hoballah, A., El Haidari, R., Badran, R., Jaber, A., Mansour, S., and Abou-Abbas, L. (2022). Smoking status and sars-cov-2 infection severity among lebanese adults: A cross-sectional study. *BMC Infectious Diseases*, 22(1):746.

Kashyap, V., Dhasmana, A., Massey, A., Kotnala, S., Zafar, N., Jaggi, M., Yallapu, M., and Chauhan, S. (2020). Smoking and covid-19: Adding fuel to the flame. *International Journal of Molecular Sciences*, 21(18):6581.

Kendall, M. and Stuart, A. (1979). *The Advanced Theory of Statistics: Inference and Relationship*. Hodder Arnold, London.

Lippi, G. and Henry, B. (2020). Active smoking is not associated with severity of coronavirus disease 2019 (covid-19). *European Journal of Internal Medicine*, 75.

Maziak, W., Nakkash, R., Bahelah, R., Husseini, A., Fanous, N., and Eissenberg, T. (2014). Tobacco in the arab world: Old and new epidemics amidst policy paralysis. *Health Policy and Planning*, 29(6):784–794.

Organization, W. H. (2021). Global report on trends in prevalence of tobacco use 2000-

2025, fourth edition.

Organization, W. H. (2023). Who coronavirus (covid-19) dashboard.

Thiel, H. (1950). A rank-invariant method of linear and polynomial regression analysis: 3. confidence regions for the parameters of polynomial regression equations. *Indignations Mathematicae*, 1(2):467–482.

Usman, M., Siddiqi, T., Khan, M., Patel, U., Shahid, I., Ahmed, J., Kalra, A., and Michos, E. (2021). Is there a smoker's paradox in covid-19? *BMJ Evidence-Based Medicine*, 26(6):279–284.

Wald, A. (1940). The fitting of straight lines if both variables are subject to error. *The Annals of Mathematical Statistics*, 11(3):284–300.