



# Analyzing Gensini Score as a Semi-Continuous Outcome

Homa Kashani, MSc<sup>1</sup>, Hojjat Zeraati, PhD<sup>1\*</sup>, Kazem Mohammad, PhD<sup>1</sup>,  
Hamidreza Goodarzynejad, MD<sup>2</sup>, Mahmood Mahmoudi, PhD<sup>1</sup>, Saeed  
Sadeghian, MD<sup>2</sup>, Mohammadali Boroumand, MD<sup>2</sup>

<sup>1</sup>Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran.

<sup>2</sup>Tehran Heart Center, Tehran University of Medical Sciences, Tehran, Iran.

Received 02 November 2015; Accepted 06 January 2016

## Abstract

**Background:** Investigators frequently encounter continuous outcomes with plenty of values clumped at zero called semi-continuous outcomes. The Gensini score, one of the most widely used scoring systems for expressing coronary angiographic results, is of this type. The aim of this study was to apply two statistical approaches based on the categorization and original scale of the Gensini score to simultaneously assess the association between covariates and the presence and severity of coronary artery disease (CAD).

**Methods:** We considered the data on 1594 individuals admitted to Tehran Heart Center with CAD symptoms from July 2004 to February 2008. The participants' baseline demographic and clinical characteristics were collected, and their coronary angiographic results were expressed through the Gensini score. The generalized ordinal threshold and two-part models were applied for the statistical analyses.

**Results:** Totally, 320 (20.1%) individuals had a Gensini score of zero. The results of neither the two-part model nor the generalized ordinal threshold model showed a significant association between Factor V Leiden and the occurrence of CAD. However, based on the two-part model, Factor V Leiden was associated with the severity of CAD, such that the Gensini score increased by moving from a wild genotype to a heterozygote ( $\beta = 0.44$ ; 95% CI: 0.20-0.69 in logarithm scale) or a homozygote mutant ( $\beta = 0.70$ ; 95% CI: 0.28- 1.12 in logarithm scale). The proportional odds assumption was not met in our data ( $\chi^2 = 54.26$ ;  $p$  value  $< 0.001$ ); however, a trend toward severe CAD was also observed at each category of the Gensini score using the generalized ordinal threshold model.

**Conclusion:** We conclude that besides loss of information by sorting a semi-continuous outcome, violation from the proportional odds assumption complicates the final decision, especially for clinicians. Therefore, more straightforward models such as the two-part model should receive more attention while analyzing such outcomes.

*J Teh Univ Heart Ctr 2016;11(2):55-61*

**This paper should be cited as:** Kashani H, Zeraati H, Mohammad K, Goodarzynejad HR, Mahmoudi M, Sadeghian S, Boroumand MA. Analyzing Gensini Score as a Semi-Continuous Outcome. *J Teh Univ Heart Ctr 2016;11(2):55-61.*

**Keywords:** Coronary artery disease • Coronary angiography • Data interpretation, statistical

\*Corresponding Author: Hojjat Zeraati, Professor of Biostatistics, Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Poursina Street, Keshavarz Boulevard, Tehran, Iran. 1417613151. Tel: +98 21 88989123. Fax: +98 21 88989127. E-mail: zeraatih@tums.ac.ir.

## Introduction

Investigators frequently confront continuous outcomes, wherein a large number of observations equal to zero. In literature, these types of variables are called “semi-continuous”, “zero-inflated continuous”, and/or “clumped at zero” data.<sup>1</sup> Examples of semi-continuous outcomes are many in medical, economic, and ecological studies. In cardiovascular medicine, the Gensini score, one of the most widely used scoring systems,<sup>2</sup> is an example where a zero value indicates no luminal stenosis within the coronary artery tree, representing patients without coronary artery disease (non-CAD group).<sup>3</sup> This is a complete and useful, but not ideal, scoring system developed by Gensini,<sup>3</sup> which emphasizes more on the severity of CAD. It takes into account the information about the geographical location and degree of luminal narrowing as well as the cumulative effect of multiple obstructions. The severity of stenosis is indicated by the reduction in lumen diameter, and a nonlinear score is assigned to each lesion based upon it. Then, according to the functional importance of the area of each lesion in the coronary tree, a multiplier is applied. The Gensini score is the sum of the lesion scores. Although the Gensini score provides a quantitative variable, it is rarely used quantitatively in statistical analyses. The reason is that it is a semi-continuous outcome with a relatively large number of observed values clustered at zero that cannot be expressed through a single distribution and the right-skewed non-zero values cannot be transformed to normality.

A simple and common practice is to recode the Gensini score into a dichotomous variable of having CAD or not or to sort it into an ordinal variable using specific cut points. Nonetheless, such an approach leads to loss of information. Alternatively, the two-part approach uses one equation (generalized linear model usually using probit or logit link function) to model the likelihood of having a non-zero value and a second equation (ordinary linear regression) to model the values greater than zero. Even though the two-part model is conceptually attractive, it was developed in econometrics in the early 1980s.<sup>4</sup> Other statistical methods on modeling semi-continuous data including Tobit<sup>5</sup> and Heckman sample selection<sup>6-8</sup> were also originally described in econometric literature. The application of these models has scarcely been tested in medical studies. Hence, we placed emphasis more statistically rather than clinically on the Gensini score as a semi-continuous outcome and applied two approaches based on the categorization and use of the Gensini score in its original scale to simultaneously assess the association between covariates and the presence and severity of CAD.

## Methods

In this cross-sectional study, we used the data of a

published research on the association between Factor V Leiden with the presence and severity of CAD. The details of the data collection procedure and the participants were previously described.<sup>9</sup> Briefly, a total of 1594 individuals with symptoms related to CAD who were admitted to Tehran Heart Center (Tehran, Iran) for elective coronary angiography between July 2004 and February 2008 were included. Coronary angiography was performed via the percutaneous femoral approach using standard angiographic techniques, and the severity of CAD was expressed with a well-known Gensini score.<sup>3</sup> The participants' baseline demographic and clinical characteristics (including age, sex, body mass index [BMI], smoking status, family history of CAD, diabetes, hypertension, hyperlipidemia, creatinine, history of renal failure, left ventricular ejection fraction [LVEF], and Factor V Leiden) were collected after obtaining written informed consent. Concisely, current smokers were those who smoked any kind of tobacco daily or quit smoking for < 1 month, and any proven CAD in a parent or sibling (under 55 and 65 years for men and women, respectively) was considered as a positive family history of CAD. The genotype analysis for Factor V Leiden was performed using the polymerase chain reaction-based restriction fragment length polymorphism (PCR-RFLP). The local ethics committee approved the study protocol.

After describing the data, we applied two approaches to analysis: the ordinal threshold model based on the categorization and the two-part model based on the original scale of the Gensini score.

As in ordinal threshold model a semi-continuous outcome is grouped into a number of ordered categories so that the first category contains zero outcomes and cut points will be selected to define the other categories,<sup>10</sup> we defined tertiles of the Gensini score as cut points and applied the logit link function. The relationship between Factor V Leiden and the Gensini score was assessed in unadjusted and adjusted models. Covariates with *p* values < 0.2 in the univariable analysis were considered in the adjusted model. Although the cumulative odds ratios (ORs) with 95% confidence intervals (CIs) were presented in the results, we found that the crucial proportional odds (PO) assumption, which investigates whether the relationship between the cumulative probabilities of the ordinal outcome categories and the covariates is the same for each category of the outcome, was not met in our data. Therefore, the generalized ordered logit model was fitted afterward, which provides different estimates at each category of the Gensini score.

Two-part model considers a semi-continuous outcome as a mixture of two parts: “occurrence or binary” and “intensity (severity) or continuous”.<sup>4</sup> With respect to our data, it means that there are two processes: one governs the occurrence of CAD (non-zero vs. zero Gensini score) and the other one manages the severity of CAD (positive values of the Gensini score) conditional on the occurrence of CAD. The logit link



function was used in the occurrence part, and the logarithm of the Gensini score was considered in the severity part to handle the skewness conditional where the non-zero Gensini score was observed. The effect of covariates with  $p$  values  $< 0.2$  in the univariable analysis at each part of the model was adjusted in a multivariable model to investigate the relationship between Factor V Leiden and the Gensini score. The effects of the covariates on the occurrence of CAD were reported using OR with 95% CI, and their effects on the severity of CAD were presented through  $\beta$  estimates with 95% CI.

Commands 'twopm', 'ologit', 'omodel', and 'gologit2' in STATA software (StataCorp. 2011. Stata Statistical Software: Release 12. College Station, TX: StataCorp LP) were used for the statistical analyses.

## Results

The median age of the 1594 participants was 58 years old (1st quartile = 51, and 3rd quartile = 66), and 1022 (64.1%) were male. The individuals' baseline characteristics are shown in Table 1. The median of the Gensini score was 30.5 (min = 0, and max = 450), and 320 (20.1%) individuals had a zero score (non-CAD group). The histogram of the Gensini score is depicted in Figure 1. The relatively high fraction of zeros and the skewness are obvious in this figure.

Table 1. Baseline characteristics of the study participants

	Mean $\pm$ SD or n (%)
Age (y)	58 (51, 66)*
Sex (male)	1022 (64.1)
BMI (kg/m <sup>2</sup> )	27.77 $\pm$ 4.50
Smoking status	
Non-smoker	941 (59.0)
Ex-smoker	306 (19.2)
Current smoker	347 (21.8)
LVEF (%)	52.49 $\pm$ 10.74
Diabetes mellitus	449 (28.2)
Hypertension	810 (50.8)
Hyperlipidemia	1044 (65.5)
Creatinine (mg/dL)	1.19 $\pm$ 0.50
Family history of CAD	341 (21.4)
History of renal failure	25 (1.6)
Factor V Leiden	
Wild	1456 (91.3)
Heterozygote	106 (6.7)
Homozygote mutant	32 (2)

\*Median (1st quartile, 3rd quartile)

BMI, Body mass index; LVEF, Left ventricular ejection fraction; CAD, Coronary artery disease

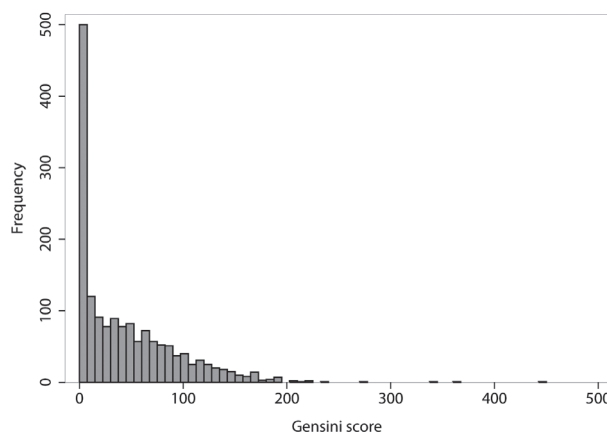


Figure 1. Histogram of the Gensini score of the study participants

Ordinal threshold model: The Gensini score cut points for the non-zero values were 24.5 (percentile 33.3) and 68 (percentile 66.7). Therefore, 0,  $> 0$  to  $\leq 24.5$ ,  $> 24.5$  to  $\leq 68$ , and  $> 68$  values were considered as Gensini score groups. The univariable cumulative logistic regression models revealed that all the covariates except hypertension ( $p$  value = 0.903) had a statistically significant relation with the Gensini score and were considered as potential confounders in the association between Factor V Leiden and the Gensini score. The results of the adjusted ordinal threshold model are shown in Table 2. However, we found that the proportional odds assumption was violated in our data ( $X^2 = 54.26$ ,  $df = 22$ ;  $p$  value  $< 0.001$ ). Therefore, the relationship between Factor V Leiden and the Gensini score was not the same in the different classes of the Gensini score and the cumulative OR for the relationship between a heterozygote and a homozygote mutant with the Gensini score relative to a wild genotype could not be considered as the unique estimates of 2.05 and 4.62, respectively.

The results of fitting the generalized ordered logit model are presented in Table 3. Obviously, due to separate estimations for the cumulative ORs at each category of the Gensini score, it is too complicated to make a general decision for the effect of the covariates. Regarding Factor V Leiden, it was difficult to conclude about the effects; however, the nonsignificant results for Gensini scores  $> 0$  versus zero Gensini score might reflect that heterozygote or homozygote mutant genotypes did not have any association with the occurrence of CAD as compared to wild genotype ( $p$  value = 0.375 and  $p$  value = 0.488) and that they merely played role in the severity of CAD. Despite the complexity of this model, the trend toward a higher Gensini score and, thus, more severe CAD with Factor V Leiden was observed by moving from a wild genotype to a heterozygote (OR = 1.81 and OR = 2.52) or a homozygote mutant (OR = 3.18 and OR = 5.87).

Table 2. Adjusted association between Factor V Leiden and Gensini score using the ordinal threshold model

	OR (95% CI)*	P Value
Factor V Leiden		
Heterozygote	2.05 (1.38-3.05)	< 0.001
Homozygote mutant	4.62 (2.06-10.35)	< 0.001
Age (y)	1.04 (1.03-1.05)	< 0.001
Sex (male)	2.23 (1.75-2.84)	< 0.001
BMI (kg/m <sup>2</sup> )	1.01 (0.98-1.03)	0.625
Smoking status		
Current smoker	1.74 (1.34-2.27)	< 0.001
Ex-smoker	1.37 (1.06-1.79)	0.018
LVEF (%)	0.95 (0.94-0.96)	< 0.001
Diabetes mellitus	1.84 (1.48-2.28)	< 0.001
Hypertension	1.56 (1.24-1.97)	< 0.001
Hyperlipidemia	1.57 (1.28-1.92)	< 0.001
Creatinine (mg/dL)	1.05 (0.82-1.35)	0.698
Family history of CAD	1.56 (1.24-1.97)	< 0.001
History of renal failure	1.41 (0.56-3.57)	0.469

\*Cumulative OR

BMI, Body mass index; LVEF, Left ventricular ejection fraction; CAD, Coronary artery disease

Two-part model: Factor V Leiden was not associated with the occurrence of CAD in the unadjusted model (p value =

0.270). Based on the univariable analysis, we found that the effect of hypertension (p value = 0.230) in the occurrence part and family history of CAD (p value = 0.363) in the severity part needed no adjustment. Table 4 shows the unadjusted and adjusted associations between Factor V Leiden and the occurrence of CAD as well as the severity of disease in the CAD group. Accordingly, Factor V Leiden only had a statistically significant effect on the severity of CAD even after adjustment for the effect of the other covariates (p value < 0.001) and not on its occurrence (OR = 1.25, 95% CI: 0.68-2.31 and OR = 1.42, 95% CI: 0.46-4.43 for heterozygote and homozygote mutant, respectively). In other words, the logarithm of the Gensini score increased by 0.44 and 0.70 by moving from a wild genotype to a heterozygote and a homozygote mutant in the CAD group, respectively.

## Discussion

In the present study, the ordinal threshold and two-part models were applied to simultaneously assess the association between Factor V Leiden and the occurrence and severity of CAD using a semi-continuous Gensini score. Using the ordinal threshold model, zero values were considered as one group and the other values were classified. We found that PO assumption was not met in our data, so separate estimates

Table 3. Adjusted association between Factor V Leiden and the Gensini score using the generalized ordered logit model

	Gensini score category					
	> 0 vs. 0		> 24.5 vs. ≤ 24.5		> 68 vs. ≤ 68	
	OR (95% CI)*	P Value	OR (95% CI)*	P Value	OR (95% CI)*	P Value
Factor V Leiden						
Heterozygote	1.31 (0.72-2.36)	0.375	1.81 (1.14-2.87)	0.011	2.52 (1.63-3.91)	< 0.001
Homozygote mutant	1.48 (0.49-4.46)	0.488	3.18 (1.29-7.82)	0.012	5.87 (2.58-13.36)	< 0.001
Age (y)	1.06 (1.06-1.07)	< 0.001	1.03 (1.02-1.05)	< 0.001	1.02 (1.01-1.04)	0.001
Sex (male)	2.79 (1.99-3.92)	< 0.001	2.20 (1.67-2.91)	< 0.001	1.69 (1.22-2.33)	0.001
BMI (kg/m <sup>2</sup> )	1.02 (0.99-1.05)	0.220	1.00 (0.97-1.02)	0.718	1.00 (0.97-1.04)	0.795
Smoking status						
Current smoker	1.81 (1.21-2.71)	0.004	2.04 (1.49-2.78)	< 0.001	1.36 (0.97-1.91)	0.071
Ex-smoker	1.57 (1.03, 2.40)	0.035	1.48 (1.09-2.02)	0.013	1.19 (0.85-1.66)	0.316
LVEF (%)	0.95 (0.94-0.97)	< 0.001	0.95 (0.94-0.97)	< 0.001	0.95 (0.94-0.96)	< 0.001
Diabetes mellitus	2.22 (1.57-3.15)	< 0.001	1.69 (1.31-2.18)	< 0.001	1.75 (1.34-2.29)	< 0.001
Hyperlipidemia	1.73 (1.30-2.31)	< 0.001	1.61 (1.27-2.05)	< 0.001	1.38 (1.05-1.81)	0.019
Creatinine (mg/dL)	1.25 (0.76-2.06)	0.386	1.00 (0.75-1.34)	0.990	1.05 (0.80-1.39)	0.722
Family history of CAD	1.92 (1.34-2.75)	< 0.001	1.38 (1.05-1.81)	0.022	1.59 (1.19-2.13)	0.002
History of renal failure	1.21 (0.10-14.71)	0.884	1.39 (0.44-4.39)	0.575	1.63 (0.60-4.42)	0.337

\*Cumulative OR

BMI, Body mass index; LVEF, Left ventricular ejection fraction; CAD, Coronary artery disease



Table 4. Association between Factor V Leiden and the Gensini score using the two-part model

Model	Variables	Occurrence of CAD		Severity of CAD	
		OR (95% CI)	P Value	$\beta$ (95% CI)*	P Value
Unadjusted					
Factor V Leiden					
	Heterozygote	1.35 (0.79-2.31)	0.267	0.49 (0.24-0.74)	< 0.001
	Homozygote mutant	1.81 (0.63-5.20)	0.271	0.89 (0.46-1.33)	< 0.001
Adjusted					
Factor V Leiden					
	Heterozygote	1.25 (0.68-2.31)	0.471	0.44 (0.20-0.69)	< 0.001
	Homozygote mutant	1.42 (0.46-4.43)	0.543	0.70 (0.28-1.12)	0.001
	Age (y)	1.06 (1.04-1.08)	< 0.001	0.01 (0.00-0.02)	0.003
	Sex (male)	2.99 (2.12-4.23)	< 0.001	0.28 (0.11-0.45)	0.001
	BMI (kg/m <sup>2</sup> )	1.02 (0.99-1.05)	0.202	-0.01 (-0.02-0.01)	0.483
Smoking status					
	Current smoker	1.72 (1.14-2.59)	0.010	0.23 (0.05-0.40)	0.010
	Ex-smoker	1.51 (0.99-2.32)	0.057	0.12 (-0.05-0.29)	0.157
	LVEF (%)	0.96 (0.94-0.97)	< 0.001	-0.02 (-0.03 - -0.02)	< 0.001
	Diabetes mellitus	2.27 (1.59-3.24)	< 0.001	0.17 (0.03- 0.31)	0.015
	Hypertension	-	-	-0.08 (-0.22-0.05)	0.239
	Hyperlipidemia	1.73 (1.29-2.32)	< 0.001	0.27 (0.13-0.41)	< 0.001
	Creatinine (mg/dL)	1.31 (0.76-2.25)	0.325	-0.02 (-0.17-0.12)	0.782
	Family history of CAD	1.87 (1.30-2.71)	0.001	-	-
	History of renal failure	1.41 (0.14- 14.41)	0.774	0.22 (-0.34-0.77)	0.441

\*For logarithm of the Gensini score

BMI, Body mass index; LVEF, Left ventricular ejection fraction; CAD, Coronary artery disease

for the aforementioned relationship in the different classes of the Gensini score were presented. Several estimates for the cumulative ORs provided a complicated model, which makes it difficult, especially for clinicians, to come to a clear conclusion. As Min and Agresti<sup>1</sup> mentioned, arbitrary selection of the cut points and the number of categories as well as loss of information due to categorizing a continuous variable are among other disadvantages of this approach. Using this model, we found it difficult to arrive at a conclusion regarding the association between Factor V Leiden and the severity of CAD since it varies in the different categories of the Gensini score. Nevertheless, a trend from a nonsignificant association with the occurrence of CAD toward a stronger association with the severity of CAD was observed by moving from a wild genotype to a heterozygote or homozygote mutant.

A more appropriate approach, the two-part model with the characteristic of considering two processes for the occurrence and severity of disease was then applied. The comprehension and interpretation of this approach is straightforward, especially for clinicians. The two-part model was preferred to its Tobit or Heckman sample selection counterparts since the Tobit model takes into account the same influence for covariates on the occurrence and severity of CAD.<sup>5</sup> However, as was observed in our data, this might not always be the

case. Also, we avoided using the Heckman sample selection model because zeros in the Gensini score are actual and not missing or censoring data.<sup>11, 12</sup> Findings based on this model clearly revealed that Factor V Leiden was not significantly associated with the occurrence of CAD neither before nor after adjustment for the effect of the other covariates. Nevertheless, when CAD happened, this factor was associated with its severity, such that as compared to wild genotype, heterozygote and homozygote mutant were associated with an increase in the Gensini score values, indicating more severe CAD. Although the trend was also shown using the generalized ordinal threshold model, it was based on the estimation of a number of ORs. It was very easily concluded through single  $\beta$  estimate from the two-part model. Another salient point is that the number of estimated ORs depends on the arbitrary selection of the number of cut points. Using these data before and considering the vessel score in analyses, Boroumand et al.<sup>9</sup> reported the existence of an association between Factor V Leiden and both the occurrence and the severity of CAD; however, using a more informative scoring system, we did not observe this relation for the occurrence of CAD. This can be explained by the difference in the classification of CAD patients. In the current study, the non-CAD group was considered as individuals with a zero Gensini score (normal subjects), whereas Boroumand et al.<sup>9</sup> considered normal and

minimal (< 50% luminal stenosis) individuals as the non-CAD group. Therefore, individuals with a small Gensini score who might be in the minimal classification of the vessel score were considered in the CAD group in the present study. Hence, it appears that one should be cautious about combining minimal coronary with normal subjects.

Chang et al.<sup>13</sup> concluded that only PO assumption played a role in selecting between the two-part and ordinal threshold models and that only in the case of PO assumption failure was it possible for the predictors of zero values to be different from those of other values; otherwise, there was no priority in choosing between these two approaches. However, in our opinion, the nature of data is very important. In our data, zero values for the Gensini score represent normal individuals without luminal stenosis, and separating them from others seems rational when investigating the effect of covariates or doing adjustment to study some favorable relationship.

In this study, we used logit and logarithm link functions for the two parts of the model because it is easy to interpret the results. The link function for the occurrence part is limited to logit or probit; however, attempts have been made to consider other link functions such as gamma or log-skew-normal for the severity part of the model.<sup>14, 15</sup> Another important issue worth noting is that since longitudinal and repeated measurements designs are commonly observed in medical studies, statisticians have recently focused more on providing suitable methods for modeling semi-continuous outcomes in these situations usually by considering random-effects in models.<sup>16-20</sup>

In this study, we focused on the Gensini score. Be that as it may, interest toward using other scoring systems such as the SYNTAX score<sup>21</sup> has been raised recently. The Gensini score is a good example of a semi-continuous outcome in clinical studies. A relatively high proportion of zero values and right skewed distribution might be frequently observed using other outcomes or scoring systems, and using proper statistical approaches to the analysis of such data will lead to more precise results.

One of the limitations of this study was that we were not able to do a comparison between the applied models easily using routine measures such as the Akaike information criterion (AIC) or Bayesian information criterion (BIC) because this requires using all data in the two models. In the two-part model, the occurrence part uses all the data and the severity part uses just the non-zero data, which leads to the estimation of the pseudo-likelihood rather than the likelihood, whereas the ordinal threshold model provides the likelihood. Therefore, we avoided this comparison, which might be studied more in further research.

## Conclusion

We conclude that besides loss of information by

sorting a semi-continuous outcome, violation from the PO assumption complicates arriving at a clear decision, especially for clinicians. Therefore, paying more attention to more straightforward models such as the two-part model is recommended when analyzing such outcomes.

## Acknowledgements

The present study was supported by Tehran University of Medical Sciences as a part of a PhD thesis.

## References

1. Min Y, Agresti A. Modeling nonnegative data with clumping at zero: a survey. *JIRSS* 2002;1:7-33.
2. Neeland IJ, Patel RS, Eshtehardi P, Dhawan S, McDaniel MC, Rab ST, Vaccarino V, Zafari AM, Samady H, Quyyumi AA. Coronary angiographic scoring systems: an evaluation of their equivalence and validity. *Am Heart J* 2012;164:547-552.
3. Gensini GG. A more meaningful scoring system for determining the severity of coronary heart disease. *Am J Cardiol* 1983;51:606.
4. Duan N, Manning WG, Morris CN, Newhouse JP. A comparison of alternative models for the demand for medical care. *J Bus Econ Stat* 1983;1:115-126.
5. Tobin J. Estimation of relationships for limited dependent variables. *Econometrica* 1958;26:24-36.
6. Heckman J. Shadow prices, market wages, and labor supply. *Econometrica* 1974;42:679-694.
7. Heckman JJ. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Ann Econ Soc Meas* 1976;5:475-492.
8. Heckman JJ. Sample selection bias as a specification error. *Econometrica* 1979;47:153-161.
9. Boroumand M, Pourgholi L, Ziaee S, Anvari MS, Jalali A, Goodarzynejad H. The association between Factor V Leiden with the presence and severity of coronary artery disease. *Clin Biochem* 2014;47:356-360.
10. Saei A, Ward J, McGilchrist CA. Threshold models in a methadone programme evaluation. *Stat Med* 1996;15:2253-2260.
11. Dow WH, Norton EC. Choosing between and interpreting the Heckit and two-part models for corner solutions. *Health Serv Outcomes Res Methodol* 2003;4:5-18.
12. Duan N, Manning WG, Morris CN, Newhouse JP. Choosing between the sample-selection model and the multi-part model. *J Bus Econ Stat* 1984;2:283-289.
13. Chang BH, Pocock S. Analyzing data with clumping at zero. An example demonstration. *J Clin Epidemiol* 2000;53:1036-1043.
14. Chai HS, Bailey KR. Use of log-skew-normal distribution in analysis of continuous data with a discrete component at zero. *Stat Med* 2008;27:3643-3655.
15. Mian RI, Hasan MT. Two-part pattern-mixture model for longitudinal incomplete semi-continuous toenail data. *Int J Stat Med Res* 2012;1:120-127.
16. Belasco E, Ghosh S. Modelling semi-continuous data using mixture regression models with an application to cattle production yields. *J Agric Sci* 2012;150:109-121.
17. Liu L, Strawderman RL, Cowen ME, Shih YC. A flexible two-part random effects model for correlated medical costs. *J Health Econ* 2010;29:110-123.
18. Olsen MK, Schafer JL. A two-part random-effects model for semi-continuous longitudinal data. *J Am Stat Assoc* 2001;96:730-745.
19. Tooze JA, Grunwald GK, Jones RH. Analysis of repeated measures data with clumping at zero. *Stat Methods Med Res* 2002;11:341-



- 355.
20. Xie H, McHugo G, Sengupta A, Clark R, Drake R. A method for analyzing longitudinal outcomes with many zeros. *Ment Health Serv Res* 2004;6:239-246.
  21. Sianos G, Morel MA, Kappetein AP, Morice MC, Colombo A, Dawkins K, van den Brand M, Van Dyck N, Russell ME, Mohr FW, Serruys PW. The SYNTAX Score: an angiographic tool grading the complexity of coronary artery disease. *EuroIntervention* 2005;1:219-227.

Archive of SID