J Biostat Epidemiol. 2016; 2(3): 118-24.

Original Article

Fuzzy risk factors in case–control studies; an example of relationship between obesity and hypertension based on National Surveillance of Risk Factors of Non-Communicable Diseases-2007 Data

Alipasha Meysamie[1], Ehsan Razzaghi[2*]

[1] Department of Community and Preventive Medicine, School of Medicine, Tehran University of Medical Sciences, Tehran, Iran
[2] Information Technology and Statistics Center, Tehran University of Medical Sciences, Tehran, Iran

ARTICLE INFO

Received 05.02.2016
Revised 06.03.2016
Accepted 14.05.2016
Published 27.08.2016

Key words:
Fuzzy risk factor;
Fuzzy set;
Case-control study;
Fuzzy odds ratio;
National Surveillance of Risk Factors of Non-Communicable Diseases

ABSTRACT

Background & Aim: There are some ambiguities in assessment of associations between continuous risk factors and different health outcomes usually from different cut points. Data loss, near to the cut point values different categorization, and no real definition of risk are important limitations for usual odds ratios (ORs). Fuzzy method considers a specific membership function for all numbers in range of the variable and can provide a similar OR. In this study, we used a large data set for these different measures calculation and making a comparison between them according to their privileges.

Methods & Materials: The study was conducted on noncommunicable diseases risk factors surveillance data set (National Surveillance of Risk Factors of Non-Communicable Diseases-2007) with regard to obesity and abdominal (central) obesity as risk factors for hypertension according to a "fuzzy risk factor" approach and usual approach based on regular cut points in different literature. OR of chances to have hypertension calculated by both methods and compared with each other.

Results: ORs with usual and fuzzy methods of calculations had similarities and some differences in amount, confidence interval and confidence length. With different cut points (for waist circumference), variation between different calculations was high. Fuzzy OR was more sensitive and resistant to minor change in individual data than the others.

Conclusion: $OR_{\{Fuzzy\}}$ measures the association of exposure to risk factors with different outcomes in a closer form of clinical reality with no dependency to any cut point selection, less variability and more resistance to data variation and can be suggested as a good estimator.

Introduction

In categorical risk factors such as gender, there will be no ambiguity for defining exposed and non-exposed groups, but in quantitative risk factors, there will be confusion in emerge. Some variables such as body mass index (BMI) and waist circumference can be used to assess obesity which is continuous and need cur points for classification. For example, usually, obesity is determined as BMI equal or more than 30.0 kg/m^2, in this case, two subjects with BMI of 30.1 and 39.8 are both considered obese whereas the level of exposure is very different in these patients. In the same way, two people with a BMI of 18.9 and 29.7 are considered as non-obese but with obviously different levels of exposure. In addition, by using just one cut point, there will an unrealistic jump for the

* Corresponding Author: Ehsan Razzaghi, Postal Address: Information Technology and Statistics Center, Tehran University of Medical Sciences, Tehran, Iran.
Email: razzaghi@farabi.tums.ac.ir

Please cite this article in press as: Meysamie A, Razzaghi E. Fuzzy risk factors in case–control studies; an example of relationship between obesity and hypertension based on National Surveillance of Risk Factors of Non-Communicable Diseases-2007 Data. *J Biostat Epidemiol. 2016; 2(3): 118-24*

values close to the cutoff point. For example, a person with a BMI of 29.9 is considered non-obese and another person with a BMI of 30.2 is considered obese, whereas there may be no real difference between them according to the clinical risk related to obesity. Even with considering several different cut points according to other variables such as gender and age, for example, what usually used according to the International Diabetes Federation (IDF) criteria for waist circumference cut points as 94.0 cm for men and 80.0 for women (1), the previously mentioned ambiguities remain sustained. When in a study there are many numeric risk factors, e.g., BMI, blood pressure and waist circumference the complicated interpretation of the level of exposure become more obvious.

If in contrast to usual categorization of the population to single 0 or 1 value, we define level of exposure to a risk factor with a range of numbers between 0 and 1, these ambiguities may be somehow resolved (2). The first method in which we define exposure to risk factors as 0 or 1, expressed in the format of an “ordinary (crisp) set,” which is mostly appropriate only for “ordinary (crisp) concepts.” The second approach which is expressed as “fuzzy concept” will not lead us to those uncertainties. Obesity, abdominal obesity, blood pressure, and many other concepts shall be routinely considered as fuzzy concepts and shall be defined based on a “fuzzy set” (2, 3). In medical science, many risk factors shall be expressed as fuzzy concept. In this study, the two different methods of ordinary and fuzzy sets compared for calculating the risk of hypertension according to obesity using the National Data of National Surveillance of Risk Factors of Non-Communicable Diseases (SuRFNCD-2007) in IR Iran.

Methods

Definition of fuzzy set for exposed subjects: The function χ_E (x) as shown in (A1) considered for the risk related to obesity in the reference population in which x is BMI, in this case, the range of values of χ_E (.) function will be {0,1} binary. According to this function all individuals with a BMI of at least 30.0, will have χ_E (.) of 1 and it is an ordinary set of obesity risk factor. In this case, we have an ordinary set of obese persons and χ_E (.) is an indicator (characteristic) function of it. Hence, χ_E (.) function is equivalent to the definition of obesity with one cutoff point of 30.0 with all ambiguities.

E(x) function is also defined as shown in (A2) as an extension of χ_E (.) but the range of values instead of {0,1} binary, is [0,1] interval. So in contrast of χ_E (.) which is an ordinary set E(.) is a fuzzy set and express membership of the obese persons into fuzzy set with value between 0 and 1. Here in E(.), we calculated the degree of membership of each individual in obese fuzzy set (2, 4) which will show the degree of exposure to obesity and can be interpreted clinically as morbidity degree to obesity. For example, the morbidity degree to obesity for two individuals with BMI 30.1 and 39.8 will be 0.508 (50.8%) and 0.996 (99.6%), respectively, which show a considerable difference between them in terms of obesity risk; In another example, morbidity degree to obesity for two people with a BMI of 29.9 and 30.2, will be 0.490 (49.0%) and 0.516 (51.6%), respectively, which show a very little difference; In this function morbidity degree to obesity for a person with a BMI of 18.9 is equal to 0.000 (0.0%) and for another person with BMI of 29.7 is 0.470 (47.0%) and according to this definition the mass function of all individuals with a BMI of < 30.0, is a maximum of 0.500 (50.0%).

$$\chi_E(x) = \begin{cases} 0 & ; \quad x < 30.0 \\ 1 & ; \quad x \geq 30.0 \end{cases} \qquad \text{(A1)}$$

$$E(x) = \begin{cases} 0 & ; \quad x < 25.0 \\ 0.1x - 2.5 & ; \quad 25.0 \leq x < 30.0 \\ 0.08x - 1.9 & ; \quad 30.0 \leq x < 35.0 \\ 0.02x + 0.2 & ; \quad 35.0 \leq x < 40.0 \\ 1 & ; \quad x \geq 40.0 \end{cases} \qquad \text{(A2)}$$

The characteristic function of the obese persons in ordinary set and membership function in fuzzy set has been shown in figure 1. In figure 2, ordinary set with crisp boundaries and fuzzy set with the turbulent boundaries compared with each other in a Venn diagram (5); the diagram shows that membership in a fuzzy set has

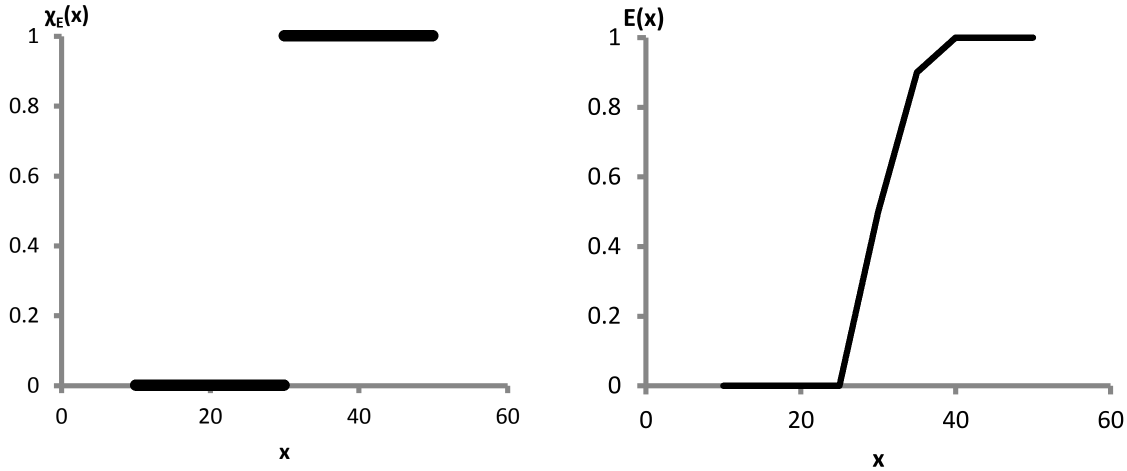

Figure 1. Characteristic function of obese persons in ordinary set (Left) and membership function of obese persons in fuzzy set (Right)

various degrees but in ordinary set there will be only two probable membership of obesity. Hence, we name the risk factor in a fuzzy concept defined by a fuzzy set as a "Fuzzy Risk Factor" to estimate the odds ratio (OR) according to fuzzy concept.

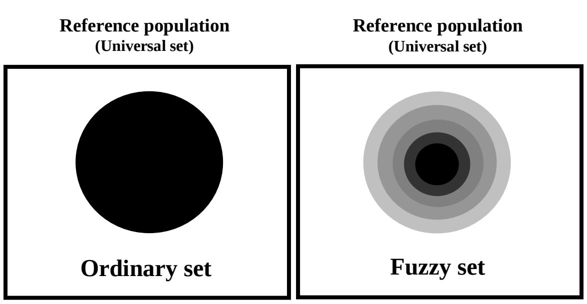

Figure 2. Ordinary sets and fuzzy sets in Venn diagram

OR calculation for fuzzy risk factors: In a case–control study, after selecting two independent samples of cases and controls (as shown in Table 1) OR can be easily calculated according to the formula (A3) in which a and c indicate the number of exposed people to a given risk factor among cases and controls, respectively, and b and d the number of non-exposed people to a given risk factor in same groups (6):

$$OR = \frac{ad}{bc} \qquad (A3)$$

Table 1. Usual pattern for data presentation in case–control studies

Group	Cases	Controls
Exposured	a	c
Unexposured	b	d
Total	m	n

In ordinary set with χ_E (.) characteristic function, each of exposed and non-exposed cases is equal to 1 and 0, respectively. Therefore, $\sum_{Cases} xE(X)$ is equal to the total number of cases exposed to risk factor which is shown with a:

$$a = \sum_{Cases} \chi_E(x) \qquad (A4)$$

Similar argument obtained for controls as c:

$$c = \sum_{Controls} \chi_E(x) \qquad (A5)$$

And b = m−a and d = n−c (A6).

In fuzzy risk factor of exposures defined by membership function of E(.) in fuzzy set since the membership function is a generalization of the characteristic function (2, 7) to determine a and c (A3) to calculate OR, term χ_E (x) in (A4) and (A5) shall be replaced with E(x) as following relations:

$$a^* = \sum_{Cases} E(x) \qquad (A7)$$

$$c^{*} = \sum_{Controls} E(x) \quad (A8)$$

Therefore, a* and c* are total degrees of exposure to fuzzy risk factor in cases and controls, respectively, and we will have b* = m−a* and d* = n−c* (A9).

Finally, with replacement of values a, b, c and d with a*, b*, c* and d* equation (A10) will be obtained to calculate the OR for a fuzzy risk factor:

$$OR_{\{Fuzzy\}} = \frac{a^{*}d^{*}}{b^{*}c^{*}} \quad (A10)$$

Eventually $OR_{\{Fuzzy\}}$ is an adjusted OR according to the level of exposures through fuzzy sets.

Like OR and by the values a, b, c and d, the asymptotic confidence intervals and asymptotic tests of significance (6) can be calculated for $OR_{\{Fuzzy\}}$ just by replacing a, b, c and d with a*, b*, c*, and d* in formulas.

Figure 3 shows the asymptotic distribution Log[$OR_{\{Fuzzy\}}$] based on 30 repetitive selections without replacement of 50 case and 50 control. As it had been shown, Log[$OR_{\{Fuzzy\}}$] distribution is similar to Log[OR] distribution, with a small nonsignificant positive skewness (Skewness = 0.24, standard error = 0.43; P = 0.5770) and is normal (P = 0.8790).

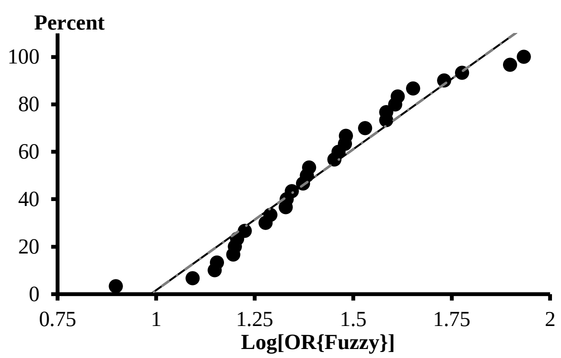

Figure 3. Asymptotic distribution of Log[$OR_{\{Fuzzy\}}$]

For assessment of the sensitive response of each measurement, we considered changes in one measurement as a test for BMI of one individual in the study, and then ORs has been recalculated and compared with previous measures.

There is no defined procedure for calculating the $OR_{\{Fuzzy\}}$ in known statistical software but different software such as SPSS, Minitab, SAS, and Excel can easily be used to perform these calculations. Hence, we used SPSS (version 20.0, IBM Corporation, Armonk, NY) for calculations and analysis according to assumptions and different criteria.

OR and $OR_{\{Fuzzy\}}$ estimation used to evaluate the associations between obesity and central obesity with hypertension in a case–control study according to the SuRFNCD-2007. The study sample consisted of 4927 residents aged 15-64 in IR Iran, including 826 cases (with hypertension) and 4101 controls (without hypertension). Details of sampling and methodology of the SuRFNCD-2007 has been discussed in other studies. Hypertension was defined as having a systolic blood pressure of at least 140.0 mmHg or a diastolic blood pressure of at least 90.0 (8). Obesity-related risk factors were calculated by measuring BMI and waist circumference for abdominal obesity according to the following criteria:

- IDF criteria for abdominal obesity: Minimum waist circumference of 94.0 for men and 80.0 for women (1)
- National Index criteria for abdominal obesity: Minimum waist circumference of 90.0 for men and women (8)
- Adult treatment panel III (ATPIII) criteria for abdominal obesity: Minimum waist circumference of 102.0 for men and 88.0 for women (8).

Calculation of the degree of exposure to abdominal obesity in a fuzzy set of risk factors: Taking into account a membership function corresponding to (A11) to define a fuzzy set of individuals with abdominal obesity based on waist circumference and abdominal obesity expose degree is calculated based on it, membership function of the fuzzy set in figure 4 has been drawn.

$$F(X) = \begin{cases} 0 & ; \quad X < 80.0 \\ 0.06X - 4.8 & ; \quad 80.0 \le X < 85.0 \\ 0.04X - 3.1 & ; \quad 85.0 \le X < 90.0 \\ 0.06X - 4.9 & ; \quad 90.0 \le X < 95.0 \\ 0.04X - 3 & ; \quad 95.0 \le X < 100.0 \\ 1 & ; \quad X \ge 100.0 \end{cases} \quad (A11)$$

Table 2. Descriptive information of studied variables

Group	Sex		Blood pressure (mmHg)		BMI (kg/m²) (Mean ± SD)	Waist circumference (cm) (Mean ± SD)
	Men (%)	Women (%)	Systolic (Mean ± SD)	Diastolic (Mean ± SD)		
Hypertensive	45.3	54.7	150.1 ± 16.9	95.8 ± 9.6	30.5 ± 4.6	99.4 ± 11.4
Normotensive	51.3	48.7	118.6 ± 15.0	76.7 ± 9.4	25.2 ± 5.2	85.3 ± 13.4
Total	50.3	49.7	123.9 ± 19.3	79.9 ± 11.8	26.0 ± 5.5	87.7 ± 14.1

SD: Standard deviation, BMI: Body mass index

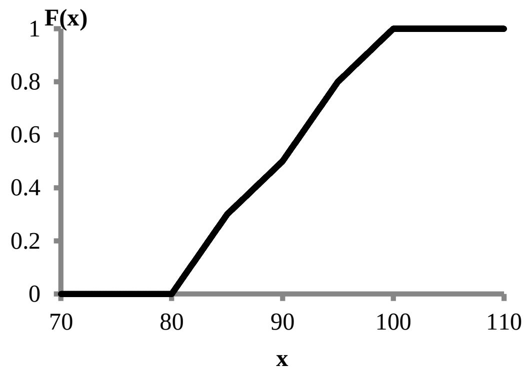

Figure 4. Fuzzy set membership function individuals with abdominal obesity

Results

Basic information about the studied variables is presented in table 2.

Different obesity risk identifications and data related to exposure distribution in different subgroups have been depicted in table 3. Different ORs has been calculated according to the different definitions of usual and fuzzy calculation for both BMI and Waist circumference of studied population show somehow different estimates and 95% confidence intervals and length of them.

By replacing only one case measure of BMI which was equal to 32.2 with 31.2, $OR_{\{Fuzzy\}}$ reduced from 4.103 to 4.102, and there was no change in usual OR. Again by replacing only one case BMI of 29.6 with 30.1, the usual OR changed from 4.554 to 4.576 but $OR_{\{Fuzzy\}}$ shows a wise change from 4.103 to 4.104.

Table 3. Usual and fuzzy risk factor OR calculation in different subjects according to hypertensive status IDF, NI and ATPIII criteria

Risk factor	Calculating method	Criteria to determine exposure	Exposed to risk factor		OR Value	P (two-tailed)	95% Confidence Interval	95% Confidence Length
Obesity	Usual (cut-off point)	Having a body mass index of at least 30.0	Hypertensive Normotensive	46.1% 15.8%	4.6	< 0.0010	(3.9, 5.3)	1.4
	Fuzzy risk factor	Determine the degree (level) of exposure from E(.) in (A2)			4.1	< 0.0010	(3.5, 4.8)	1.3
Abdominal obesity	Usual (cut-off point)	IDF: Minimum of 94.0 for men and 80.0 for women waist	Hypertensive Normotensive	86.9% 43.7%	8.6	< 0.0010	(6.9, 10.6)	3.7
		NI: Minimum of 90.0 for men and women waist	Hypertensive Normotensive	83.3% 37.2%	8.4	< 0.0010	(6.9, 10.2)	3.3
		ATPIII: Minimum of 102.0 for men and 88.0 for women waist	Hypertensive Normotensive	64.5% 25.9%	5.2	< 0.0010	(4.4, 6.1)	1.7
	Fuzzy risk factor	Determine the degree (level) of exposure from F(.) in (A11)			6.1	< 0.0010	(5.1, 7.3)	2.2

OR: Odds ratio, IDF: International Diabetes Federation, NI: National Index, ATPIII: Adult treatment Panel III

Discussion

According to the findings of this study, there is no significance difference between fuzzy and usual ORs according to BMI, which is also correct about the confidence interval length. However, the relationship between abdominal obesity and hypertension shows a significant difference between different methods of calculation. The modest amount is related to fuzzy method.

Different results in different populations according to the usual ORs calculated, e.g., in a cross-sectional study in 2009 on 843 African-American adult women (mean age 53.8 years), those with waist circumference ≥ 88 had 7.2-fold increase in risk of having hypertension comparing to those with waist circumference ≤ 80 (9) and in another case–control study, conducted from 2007 to 2010 on 11,145 Americans over 18-year-old, abdominal obesity (based on ATPIII criteria) had a great effect on hypertension with OR of 3.2 (10) we can see some differences between the results of these studies and our study which can be explained somehow by the differences related to the distribution of obesity in different populations rather than differences in association of obesity and hypertension.

According to the calculation methods, fuzzy ORs seem to be more reliable and sensitive than others. As we showed with data replacement of only one individual, when the amount is less than cut-off (or more than cutoff which has not been performed in this study) a good sensitivity of fuzzy OR can be proved in contrast of usual method of OR calculation. Its importance can be emphasized according to the large sample of 826 cases and 4101 controls in our study. Besides, slight increase in BMI which was less than cutoff point to an amount which is now more than cutoff point will result in an unrealistic jump in the usual OR (especially in small studies) which is not expected for $OR_{\{Fuzzy\}}$. As we showed with replacing a case of 29.6 BMI with 30.1, usual OR changes from 4.554 to 4.576 but $OR_{\{Fuzzy\}}$ only changes from 4.103 to 4.104. So we can conclude more resistance and sensitivity of fuzzy OR in comparison of usual ORs to minor changes in data. The reason can be explained according to the close relationship between obesity definition which has been used in $OR_{\{Fuzzy\}}$ calculation, and real association between obesity and hypertension in clinical medicine.

Data lose because of cut point assumption in $OR_{\{Fuzzy\}}$ is almost nothing rather than the large amount of data lose in usual method of OR calculation. For example in usual OR calculation, we shall consider all individuals with BMI ≥ 30.0 (or < 30.0) as a single value, those will be considered in $OR_{\{Fuzzy\}}$, at least for subjects with BMI of 25.0-40.0, as their own values in calculation.

Another privilege of $OR_{\{Fuzzy\}}$ is related to small sample studies in which values of a, b, c or d may be very small or even zero which cause incalculable usual OR (equal to 0 or ∞). Although in these situations, estimator OR definition has been suggested (A12) to estimate OR (6), but in this calculation, the OR can be increased or decreased unrealistically. $OR_{\{Fuzzy\}}$ is a more probable solution for OR calculation in small studies. With normal distribution of BMI, the probability of calculation of $OR_{\{Fuzzy\}}$ and usual OR in a study with only 5 case and 5 control, will be 0.97 and 0.58 respectively, and for a study with 10 case and 10 control, these probabilities will be 0.9992 and 0.8500, respectively. Naturally, the probabilities come close to each other with increasing sample size. Since the calculation of $OR_{\{Fuzzy\}}$ even in very small samples, are very probable, adding value 0.5 to the a*, b*, c* and d* in equation (A12) will not be necessary.

$$OR = \frac{(a+0.5)(d+0.5)}{(b+0.5)(c+0.5)}$$

Limitations: Different cut points will results in different ORs in usual OR calculation. $OR_{\{Fuzzy\}}$ is independent of any cut point and this limitation cannot be considered as a limitation in this form of calculation. However, the function which shall be used for fuzzy function may be varied according to the selected membership function and may cause some variation in $OR_{\{Fuzzy\}}$ results. This effect has not been assessed in this study.

Conclusion

In this study, we used probability theory for definitions and calculating functions, in another study the aspect of uncertainty has been based on the possibility theory for fuzzy OR calculation (11). It seems that the concepts of $OR_{\{Fuzzy\}}$ estimators based on both theories are the same and those based on probability theory are simpler and more practical than those based on the possibility theory.

$OR_{\{Fuzzy\}}$ measures the association of exposure to risk factors with different outcomes in a closer form of clinical reality with no dependency to any cut point selection, less variability and more resistance to data variation and can be suggested as a good estimator.

Acknowledgments

In this study, data from National SuRFNCD-2007 used for analysis, which has been performed in the supervision of Ministry of Health, Treatment and Education of IR Iran. There was no financial support for this study.

References

1. International Diabetes Federation. The IDF consensus worldwide definition of the metabolic syndrome [Online]. [cited 2006]; Available from: URL: http://www.idf.org/webdata/docs/IDF_Meta_def_final.pdf
2. Zadeh LA. Fuzzy sets. Inform Contr 1965; 8(1): 338-53.
3. Phuong NH, Kreinovich V. Fuzzy logic and its applications in medicine. Int J Med Inform 2001; 62(2-3): 165-73.
4. Sadegh-Zadeh K. Advances in fuzzy theory. Artif Intell Med 1999; 15(3): 309-23.
5. Wikipedia. Venn diagram [Online]. [cited 2015 Aug 22]; Available from: URL: https://en.wikipedia.org/wiki/Venn_diagram
6. Agresti A. An introduction to categorical data analysis. 2nd ed. Hoboken, NJ: Wiley-Interscience; 2007.
7. Steimann F. Fuzzy set theory in medicine. Artif Intell Med 1997; 11(1): 1-7.
8. Esteghamati A, Meysamie A, Khalilzadeh O, Rashidi A, Haghazali M, Asgari F, et al. Third national Surveillance of Risk Factors of Non-Communicable Diseases (SuRFNCD-2007) in Iran: methods and results on prevalence of diabetes, hypertension, obesity, central obesity, and dyslipidemia. BMC Public Health 2009; 9: 167.
9. Warren TY, Wilcox S, Dowda M, Baruth M. Independent association of waist circumference with hypertension and diabetes in African American women, South Carolina, 2007-2009. Prev Chronic Dis 2012; 9: E105.
10. Ostchega Y, Hughes JP, Terry A, Fakhouri TH, Miller I. Abdominal obesity, body mass index, and hypertension in US adults: NHANES 2007-2010. Am J Hypertens 2012; 25(12): 1271-8.
11. Massad E, Ortega NR, Struchiner CJ, Burattini MN. Fuzzy epidemics. Artif Intell Med 2003; 29(3): 241-59.