



ELSEVIER

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

SCEW: A Microsoft Excel add-in for easy creation of survival curves

Haseeb Ahmad Khan*

Department of Biochemistry, College of Science, Bld 5, Rm 2A68, King Saud University, P.O. Box 2455, Riyadh 11451, Saudi Arabia

ARTICLE INFO

Article history:

Received 9 March 2006

Received in revised form 8 May 2006

Accepted 14 May 2006

Keywords:

Survival curves

Microsoft Excel

Algorithm

Add-in program

ABSTRACT

Survival curves are frequently used for reporting survival or mortality outcomes of experimental pharmacological/toxicological studies and of clinical trials. Microsoft Excel is a simple and widely used tool for creation of numerous types of graphic presentations however it is difficult to create step-wise survival curves in Excel. Considering the familiarity of clinicians and biomedical scientists with Excel, an algorithm survival curves in Excel worksheet (SCEW) has been developed for easy creation of survival curves directly in Excel worksheets. The algorithm has been integrated in the form of Excel add-in for easy installation and usage. The program is based on modification of frequency data for binary break-up using the spreadsheet formula functions whereas a macro subroutine automates the creation of survival curves. The advantages of this program are simple data input, minimal procedural steps and the creation of survival curves in the familiar confines of Excel.

© 2006 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Microsoft Excel is a highly versatile and perhaps the most extensively used computer package available to clinicians and biomedical scientists for a wide range of applications. Excel has been applied for clinical data management and analysis [1–3] including the data retrieval from remote clinics to central database hub [4,5]. Brown [6–8] has used Excel platform for developing protocols for biological systems including the electrophysiology of nervous system. The capability of Excel iteration has also been used to simulate the cardiac action potential to study the properties of ionic currents and their impact on heart cells [9]. Excel macros have been applied for searching potential phosphorylation sites in proteins [10] and for identification of specific sites of protein–DNA interaction [11]. An Excel tool has been used for the validation of housekeeping genes and pair-wise correlations of differentially expressed target genes [12]. Geeraerd et al. [13] have developed an Excel add-in for testing different types of microbial survival models with a direct implication with food

industry. Excel based tools have simplified the task complexities pertaining to buffer preparations [14], enzyme kinetics [15], pharmacokinetics [16,17], polymerase chain reaction [18], mutation detection [19], paternity testing [20] and microarray data analysis [21–25]. The above literature clearly supports the notion that Excel is a convenient and multipurpose tool with its excellent features of spreadsheet, graphics, database and statistical computations.

Kaplan Meier survival curves with log rank [26–28] or Cox regression [29–31] analyses are commonly used for the interpretation of survival, mortality and relapses data. Due to their inherent clarity and displaying power, curves are preferred over tables for presenting information on survival, mortality or morbidity [32]. Survival data are best displayed using a Kaplan Meier survival plot, which is basically a step-wise line graph starting with the shortest survival time (maximum survival frequency at the beginning); the subsequent down-steps describe the time course of event(s) among censored observations [33]. Although Excel offers a wide range of graphical applications including line graph, scatter graph, bar graph

* Tel.: +966 1 4675944.

E-mail address: khan.haseeb@yahoo.com.

Table 1 – Configuration of binary data points for each event (refer to Appendix A for method)

Time		Survival frequency	
Original	Configured	Original	Configured
t_1	t_1 t_1	f_1	f_1 f_1
t_n	t_n t_n	f_n	f_{n-1} f_n

Where 't' is time and 'f' survival frequency; subscripts '1' and 'n' represent the first and subsequent data entries across the columns.

and pie chart it lacks the facility of survival curves. In the past, Excel has been used for developing programs for making curves with special features [34–36]. This study reports an Excel add-in, survival curves in Excel worksheet (SCEW), for creation of step-wise survival curves directly in Excel worksheet.

2. Computational methods and theory

The primary step for creating survival curves in Excel worksheet is to convert each survival frequency into two configured data points. Earlier we had utilized the same basic idea for manual creation of survival curves using the real experimental data [37,38]. Table 1 shows the procedure for data modification for binary break-ups in order to get two data points for each event. The formula validation feature of spreadsheet is utilized to automate the modification of survival and time data followed by the creation of survival curve using an optimized macro subroutine. The representative original data input as well as their corresponding modified data which is used for

the creation of survival curves are shown in Fig. 1. For a more clear understanding of data modification, consider an example of time 16 (row 10, Fig. 1(a)), the binary data for time and survival frequencies will be as follows: time $\{t_{10}, t_{10}: 16, 16\}$, group 1 $\{f_9, f_{10}: 94, 93\}$, group 2 $\{f_9, f_{10}: 70, 60\}$ and group 3 $\{f_9, f_{10}: 45, 45\}$ (Fig. 1(b)). The total number of rows therefore doubled after the data modification.

The basic difference between the graphical output using unmodified (original data input) or modified data is shown in Fig. 2. Although both the curves have been created using the same original data (Fig. 1(a)) and strictly the same data entry procedure, Fig. 2(a) shows a simple line graph due to one data point for each frequency whereas Fig. 2(b) shows a peculiar survival curve owing to the two data points per entry (created by SCEW).

3. Program description

3.1. Software design

Microsoft Excel platform was used to develop this software due to Excel's flexibility, universal availability, excellent graphical support and macro-based automation. Although the configuration of data for binary break-up (Table 1) can be performed manually this could rather be a complicated and tedious job especially for repeated applications. On the other hand, SCEW utilizes Excel's worksheet formula function (Appendix A) for instant data configuration. The modified data are then used to construct survival curve with the help of a macro subroutine (Appendix B) and the output is automatically routed to a new Excel worksheet. This automation renders the entire methodology more convenient, faster and error-free as compared to manual procedure.

(a) Original data					(b) Modified data				
	A	B	C	D		A	B	C	D
1	Time	Group 1	Group 2	Group 3	1	Time	Group 1	Group 2	Group 3
2	0	100	100	100	2	0	100	100	100
3	2	100	100	90	3	0	100	100	100
4	4	100	95	90	4	2	100	100	100
5	6	100	81	70	5	2	100	100	90
6	8	97	81	70	6	4	100	100	90
7	10	97	81	60	7	4	100	95	90
8	12	97	75	60	8	6	100	95	90
9	14	94	70	45	9	6	100	81	70
10	16	93	60	45	10	8	100	81	70
11	18	92	60	30	11	8	97	81	70
12	20	91	60	30	12	10	97	81	70
					13	10	97	81	60
					14	12	97	81	60
					15	12	97	75	60
					16	14	97	75	60
					17	14	94	70	45
					18	16	94	70	45
					19	16	93	60	45
					20	18	93	60	45
					21	18	92	60	30
					22	20	92	60	30
					23	20	91	60	30

Fig. 1 – Comparative view of original data (as entered by the user) and modified data (instantly computed by using the spreadsheet formula function of Excel, Appendix A).

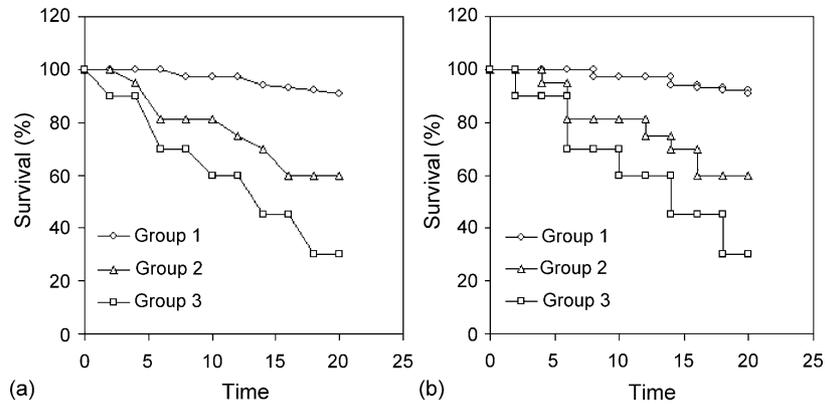


Fig. 2 – Comparative view of survival curves created by using (a) MS-Excel and (b) Excel equipped with SCEW add-in.

3.2. Installation of SCEW add-in

Open the Excel program and insert the program CD in drive (if downloading from CD, else use the appropriate drive). In the 'Tool' menu of Excel workbook, click on 'Add-Ins' and then click on 'Browse'. Locate your CD drive (or the relevant path to locate the file) and double click on 'SCEW Install'. The message "Copy SCEW to ..." will appear, click 'Yes'; the appearance of 'Survival Curve' on the menu bar indicates the proper installation of the add-in. Remove the software CD. For un-installation, follow all the above steps except selecting 'SCEW Uninstall' instead of 'SCEW Install'; the removal of 'Survival Curve' from the menu bar will indicate the proper un-installation of the add-in.

3.3. Creation of survival curves

For the creation of survival curve, the time increments and respective survival frequencies are entered in the Excel worksheet as shown in Fig. 3. The top row occupies the name of groups and considered as 'header row'. Once the data entry has been completed, click the 'Survival Curve' button in the menu bar to activate the 'Input' window. Select the range of data including the header row as shown in Fig. 3. Now clicking the 'OK' button creates the survival curve (Fig. 4) that is displayed in a new worksheet. The users can easily modify the appearance of survival curve according to their own choice for size, color, drawing, text, font, etc. using the built-in features of Excel (Fig. 5).

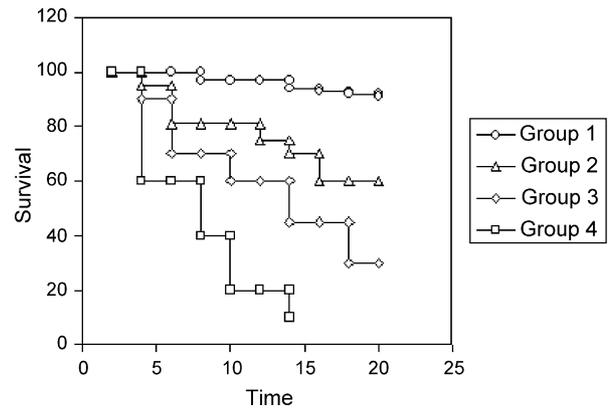


Fig. 4 – Survival curve created by SCEW program using the data shown in Fig. 3.

4. Samples of program runs

The clinical scientists usually prefer to enter the summary of different variables rather than the individual questionnaire-based data entry, particularly for small data sets which are quite often the case of various experimental pharmacological and toxicological studies [37,38]. The SCEW program is specially designed for such requirements and has been validated for proper functioning. The representative frequencies of time and survival data for different number of groups

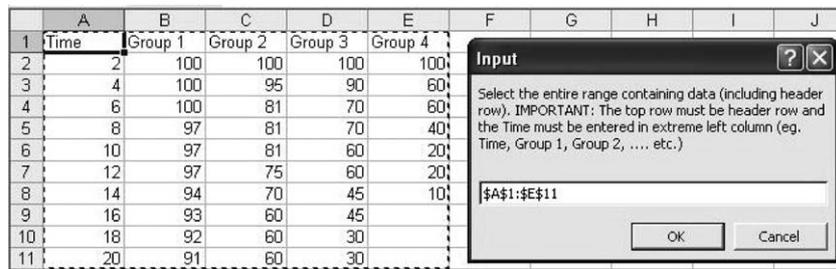


Fig. 3 – Excel worksheet showing the methodology of data entry and program execution. After data entry, click 'Survival Curve' on the menu bar to prompt the 'Input' window. Select the data range of interest and then click the 'OK' button to create survival curve.

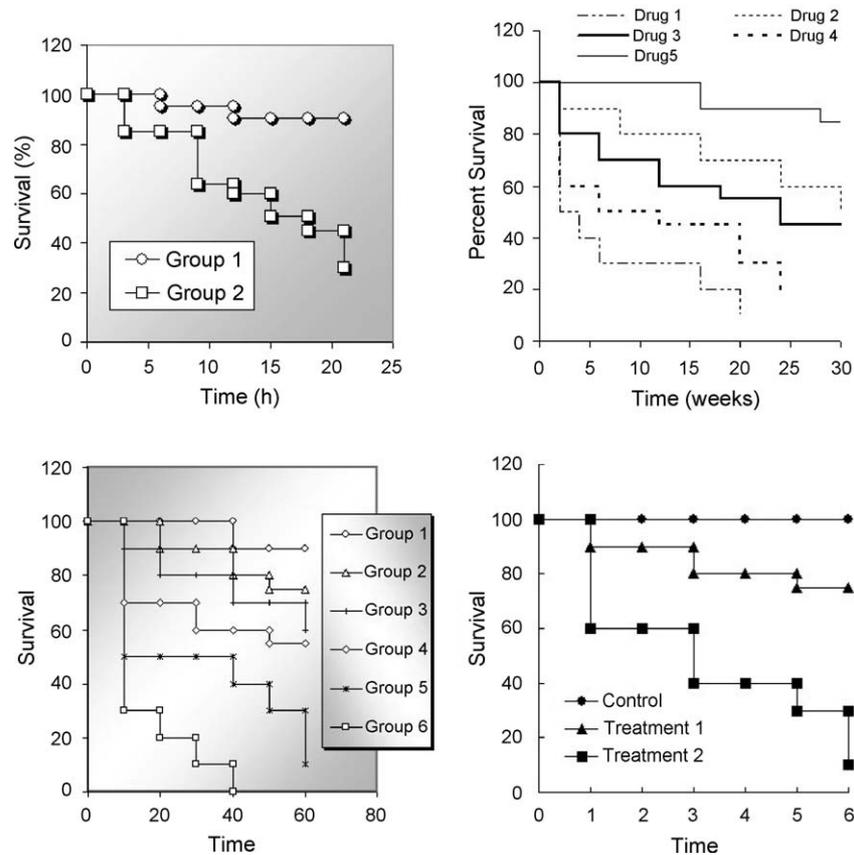


Fig. 5 – Creation of survival curves by SCEW and their post-editing for user's defined fonts, shading, line styles, etc. using the Excel's features.

were used to testify the software (Fig. 5). The software was also validated for the successful transfer of survival curves to other Microsoft Excel worksheets, Microsoft Word document and Microsoft Power point presentation using the 'copy/paste' procedure.

Although some statistical programs such as SPSS can efficiently create survival curves the methodology involved may be complex to inexperienced users. Moreover, the survival curves created in SPSS tend to be less flexible for sizing and other modifications as compared to Excel-based curves. The former curves also become refractory to further modifications if transferred to other application programs like MS-Word or MS-Power point whereas the survival curves created by SCEW are fully compatible with Windows-based programs. Moreover, Excel-based application programs are user-friendly, easily understood with less mathematical sophistication and operate within the familiar confines of Excel. The simplicity of Excel is further evidenced by its preferred use as a tool for teaching statistics to students [39]. Recently, Excel has been recognized as a poor man's program owing to its excellent multiple features at affordable price [40]. On the other hand, commercial specialist software tend to be complex, difficult to grasp for inexperienced users, and expensive.

In conclusion, the algorithm SCEW is able to enhance the utility of Excel with a novel support to create survival curves directly in Excel worksheets.

5. Hardware and software specification

A Pentium 4 computer (256 Mb RAM) running under Windows XP operating system was used. The algorithm was developed using Microsoft Excel's macro language in Excel 2003 platform. The files sizes of SCEW-Install and SCEW-Uninstall are 393 KB (84 KB zip file) and 17 KB (5 KB zip file), respectively.

6. Mode of availability of the program

The software can be obtained from the author by sending a letter of request or it can be downloaded from the URL, <http://www.cos.ksu.edu.sa/members/haseeb/>. If the software is obtained as a zip file it should be unzipped before installation.

Acknowledgement

This study was financially supported by Microsoft Corporation, USA in the form of Smart Clients for eScience Award (no. 14613) to the author.

Appendix A. Spreadsheet formula for data modification

Time column	Survival frequency column(s)
CA2 = IF(A2= "", "", A2)	CB2 = IF(B2= "", "", B2)
CA3 = CA2	CB3 = IF(B2= "", "", B2)
CA4 = IF(A3= "", "", A3)	CB4 = IF(B2= "", "", B2)
CA5 = CA4	CB5 = IF(CB4= "", "", IF(B3<>B2, B3, B2))
CA6 = IF(A4= "", "", A4)	CB6 = IF(CB5= "", "", IF(CB5 = 0, "", IF(B4= "", "", CB5)))
—continue as above for	CB7 = IF(CB6= "", "", IF(B4<>B3, B4, B3))
CA7 downwards	CB8 = IF(CB7= "", "", IF(CB7 = 0, "", IF(B5= "", "", CB7)))
	CB9 = IF(CB8= "", "", IF(B5<>B4, B5, B4))
	CB10 = IF(CB9= "", "", IF(CB9 = 0, "", IF(B6= "", "", CB9)))
	—continue as above for the remaining cells.

The top row of the worksheet is reserved for heading. The time data are entered in column 'A' starting from the cell 'A2'. Columns 'B' to 'BZ' (a capacity of holding 51 groups) are used for survival data. Column 'CA' is used for the modification (binary break-up) of time intervals whereas columns 'CB' onwards are used for the modification of survival data of different groups.

Appendix B. Main macro for the creation of survival curve

```
Sub scw()
  With mydisplaysheet
    Charts.Add
    ActiveChart.ChartType = xlXYScatterLines
    ActiveChart.SetSourceData Source:=.Range(.Cells(
    (1, 1), .Cells(1, 1)), PlotBy:=xlColumns
    u=2
    For n=1 To horizon-1
      z=Application.WorksheetFunction.Count(mydis-
      playsheet1.Range(.Cells(1, n+1),.Cells_(202, n+1)))
      + 1
      Set r1=.Range(.Cells(2, 1), .Cells(z, 1))
      Set r2=.Range(.Cells(2, u), .Cells(z, u))
      Set r3=.Range(.Cells(1, u), .Cells(1, u))
      ActiveChart.SeriesCollection(n).XValues = r1
      ActiveChart.SeriesCollection(n).Values = r2
      ActiveChart.SeriesCollection(n).Name = r3
      u = u + 1
      ActiveChart.SeriesCollection.NewSeries
    Next n
  End With
  ActiveChart.Location Where:=xlLocationAsObject,
  Name:="Survival Curve"
End Sub
```

The above code applies for the main subroutine used for the automation of survival curves. Additional codes were used for controlling the input box and other formatting features.

REFERENCES

- [1] P.P. Gomes, L.A. Passeri, J.R. Barbosa, A 5-year retrospective study of zygomatico-orbital complex and zygomatic arch fractures in Sao Paulo State, Brazil, *J. Oral Maxillofac. Surg.* 64 (2006) 63–67.
- [2] D.R. Scribner, J.L. Walker, G.A. Johnson, S.D. McMeekin, M.A. Gold, R.S. Mannel, Laproscopic pelvic and paraaortic lymph node dissection: analysis of the first 100 cases, *Gynecol. Oncol.* 82 (2001) 498–503.
- [3] P.S. Craighead, K. Sait, G.C. Stuart, K. Arthur, J. Nation, M. Duggan, D. Guo, Management of aggressive histologic variants of endometrial carcinoma at the Tom Baker Cancer Center between 1984 and 1994, *Gynecol. Oncol.* 77 (2000) 248–253.
- [4] R. Achuthan, K. Grover, F. MacFie, A critical evaluation of the electronic surgical logbook, *BMC Med. Edu.* 6 (2006) 15.
- [5] R. Glazebrook, B. Chater, P. Graham, et al., Evaluation of the ACRRM national radiology program for Australian rural and remote medical practitioners, *Rural Remote Health* 5 (2005) 349.
- [6] A.M. Brown, Simulation of axonal excitability using a Spreadsheet template created in Microsoft Excel, *Comput. Meth. Prog. Biomed.* 63 (2000) 47–54.
- [7] A.M. Brown, A methodology for simulating biological systems using Microsoft Excel, *Comput. Meth. Prog. Biomed.* 58 (1999) 181–190.
- [8] A.M. Brown, A non-linear regression analysis program for describing electrophysiological data with multiple functions using Microsoft Excel, *Comput. Meth. Prog. Biomed.* 82 (2006) 51–57.
- [9] S.N. Wu, Simulations of the cardiac action potential based on the Hodgkin–Huxley kinetics with the use of Microsoft Excel spreadsheets, *Chin. J. Physiol.* 47 (2004) 15–22.
- [10] S. Wera, An EXCEL-based method to search for potential Ser/Thr-phosphorylation sites in proteins, *Comput. Meth. Prog. Biomed.* 58 (1999) 65–68.
- [11] M.J. Buck, A.B. Nobel, J.D. Lieb, ChIPOTle: a user-friendly tool for the analysis of ChIP-chip data, *Genom. Biol.* 6 (2005) R97.
- [12] M.W. Pfaffl, A. Tichopad, C. Prgomet, T.P. Neuvians, Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper-Excel-based tool using pair-wise correlations, *Biotechnol. Lett.* 26 (2004) 509–515.
- [13] A.H. Geeraerd, V.P. Valdramidis, J.F. Van Impe, GInaFiT, a freeware tool to assess non-log-linear microbial survivor curves, *Int. J. Food. Microbiol.* 102 (2005) 95–105.
- [14] H. Okamoto, K. Mori, K. Ohtsuka, H. Ohuchi, H. Ishii, Theory and computer programs for calculating solution pH, buffer formula and buffer capacity for multiple component system at a given ionic strength and temperature, *Pharm. Res.* 14 (1997) 299–302.
- [15] P.R. Bianco, G.M. Weinstock, Automated determination of beta-galactosidase specific activity, *Biotechniques* 17 (1994) 974–980.
- [16] H. Sato, S. Sato, Y.M. Wang, I. Horikoshi, Add-in macros for rapid and versatile calculation of non-compartmental pharmacokinetic parameters on Microsoft Excel spreadsheets, *Comput. Meth. Prog. Biomed.* 50 (1996) 43–52.

- [17] C. Dansirikul, M. Choi, S.B. Duffull, Estimation of pharmacokinetic parameters from non-compartmental variables using Microsoft Excel, *Comput. Biol. Med.* 35 (2005) 389-403.
- [18] R.P. Stowe, D.L. Pierson, Spreadsheet macro for setting up PCR assay tubes, *Biotechniques* 20 (1996) 1088-1089.
- [19] B.T. Messmer, SAIVGeM: spreadsheet analysis of immunoglobulin VH gene mutations, *Biotechniques* 39 (2005) 353-358.
- [20] Y. Aoki, M. Hashiyada, A. Morioka, M. Nata, K. Sagisaka, Spreadsheets of a conventional application software for calculation of plausibility of paternity: application to parentage testing with highly polymorphic markers in deceased party, *Nippon. Hoigaku. Zasshi.* 51 (1997) 196-204.
- [21] H.A. Khan, ArrayVigil: a methodology for statistical comparison of gene signatures using segregated-one-tailed (SOT) Wilcoxon signed-rank test, *J. Mol. Biol.* 345 (2005) 645-649.
- [22] H.A. Khan, ArraySolver: an algorithm for color-coded graphical display and Wilcoxon signed-rank statistics for comparing microarray gene expression data, *Comput. Func. Genom.* 5 (2004) 39-47.
- [23] R. Anbazhagan, Microarray data assembler, *Bioinformatics* 19 (2003) 157-158.
- [24] J.J. Schageman, M. Basit, T.D. Gallardo, H.R. Garner, R.V. Shohet, MarC-V: a spreadsheet-based tool for analysis, normalization, and visualization of single cDNA microarray experiments, *Biotechniques* 32 (2002) 338-344.
- [25] T. Conway, B. Kraus, D.L. Tucker, D.J. Smalley, A.F. Dorman, L. McKibben, DNA array analysis in a Microsoft Windows environment, *Biotechniques* 32 (2002) 110-119.
- [26] N. Taniai, H. Yoshida, Y. Mamada, Y. Kawano, Y. Mizuguchi, K. Akimaru, T. Tajiri, Is intraoperative adjuvant therapy effective for satellite lesions in patients undergoing reduction surgery for advanced hepatocellular carcinoma? *Hepatogastroenterology* 53 (2006) 258-261.
- [27] S.A. Shah, R. Haddad, W. Al-Sukhni, et al., Surgical resection of hepatic and pulmonary metastases from colorectal carcinoma, *J. Am. Coll. Surg.* 202 (2006) 468-475.
- [28] A. Bufalari, G. Giustozzi, M.F. Burattini, S. Servili, C. Bussotti, E. Lucaroni, E. Ricci, F. Sciannameo, Rectal cancer surgery in the elderly: a multivariate analysis of outcome risk factors, *J. Surg. Oncol.* 93 (2006) 173-180.
- [29] G.P. Rossi, G. Maiolino, T.M. Seccia, et al., Hyperhomocysteinemia predicts total and cardiovascular mortality in high-risk women, *J. Hypertens.* 24 (2006) 851-859.
- [30] B.J. Lee, S.G. Wang, J.S. Choi, J.C. Lee, E.K. Goh, M.G. Kim, The prognostic value of telomerase expression in peripheral blood mononuclear cells of head and neck cancer patients, *Am. J. Clin. Oncol.* 29 (2006) 163-167.
- [31] T. Delaunoit, P.J. Limburg, R.M. Goldberg, J.F. Lymp, E.V. Loftus Jr., Colorectal cancer prognosis among patients with inflammatory bowel disease, *Clin. Gastroenterol. Hepatol.* 4 (2006) 335-342.
- [32] E.L. Ferriman, M.A. Doulah, N.A. Simpson, Prediction of survival for preterm births, curve presents information on survival and morbidity more clearly than tables, *B.M.J.* 320 (2000) 647-648.
- [33] M.J. Campbell, D. Machin, *Medical Statistics: A Common Sense Approach*, second ed., John Wiley & Sons, UK, 1996, pp. 112-113.
- [34] R. Ward, J. Schlenker, G.S. Anderson, Simple method for developing percentile growth curves for height and weight, *Am. J. Phys. Anthropol.* 116 (2001) 246-250.
- [35] G. Gorog, An Excel program for calculating and plotting receiver operator characteristic (ROC) curves, histograms and descriptive statistics, *Comput. Biol. Med.* 24 (1994) 167-169.
- [36] M. Sakamoto, K. Akazawa, T. Kamakura, N. Kinukawa, Y. Nishioka, Y. Nose, Microsoft Excel program for creating attractive survival curves, *J. Med. Sys.* 18 (1994) 241-249.
- [37] K. Al Moutaery, S. Al Deeb, H.A. Khan, M. Tariq, Caffeine impairs neurological recovery following concussive head injury in rats, *Neurosurgery* 53 (2003) 704-712.
- [38] M. Tariq, S. Jacobs, A. Al Moutaery, M. Arshaduddin, H.A. Khan, D.A. Price Evans, Fluconazole attenuates lung injury and mortality in rat peritonitis model, *Intensive Care Med.* 29 (2003) 2043-2049.
- [39] J. Bukac, Our statistical programs in Excel, *Acta Med.* 47 (2004) 123-124.
- [40] J.P. Sewell, Getting the most from your software: using Excel as the poor man's database, *Comput. Inform. Nurs.* 24 (2006) 13-17.