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

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.

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

	A	B	C	D
1	Time	Group 1	Group 2	Group 3
2	0	100	100	100
3	2	100	100	90
4	4	100	95	90
5	6	100	81	70
6	8	97	81	70
7	10	97	81	60
8	12	97	75	60
9	14	94	70	45
10	16	93	60	45
11	18	92	60	30
12	20	91	60	30

(b) Modified data

	A	B	C	D
1	Time	Group 1	Group 2	Group 3
2	0	100	100	100
3	0	100	100	100
4	2	100	100	100
5	2	100	100	90
6	4	100	100	90
7	4	100	95	90
8	6	100	95	90
9	6	100	81	70
10	8	100	81	70
11	8	97	81	70
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).