

# Creating Survival Curves in Excel: A Step-by-Step Guide

Authored by  
**Mohammed loot**

November 9, 2025

## RECOMMENDED CITATION

Mohammed loot (2025). *Creating Survival Curves in Excel: A Step-by-Step Guide*.  
PSYCHOLOGICAL STATISTICS. Retrieved from  
<https://statistics.arabpsychology.com/?p=14386>

The [survival curve](#) is a cornerstone visualization tool utilized extensively across fields such as [biostatistics](#), epidemiology, and reliability engineering. This graphical representation offers essential insight by depicting the estimated proportion of a group--be it patients, mechanical components, or a defined population--that remains unaffected by a specific event over time. The event might represent mortality, disease recurrence, or system failure. By illustrating the probability of survival beyond any given point in time, the survival curve provides critical data for prognosis assessment and the comparative analysis of different intervention groups.

While professional statistical packages are typically employed for complex [survival analysis](#), Microsoft Excel possesses the necessary computational and charting capabilities to effectively construct a basic, yet statistically sound, survival curve using the principles of the [Kaplan-Meier estimator](#). This detailed tutorial guides you through the methodical steps required, from initial data organization to the final visualization, demonstrating how to leverage Excel for this powerful analytical task.

## Preparing the Raw Dataset for Survival Analysis

Before initiating any calculation, the raw data must be structured explicitly for survival analysis. A standard survival dataset requires two primary variables: the observation time and the event status. The time variable records the duration each subject was observed (e.g., months, years, hours) until the event occurred or the study ended. The event status is a binary indicator: typically, '1' signifies that the event of interest occurred (e.g., death, failure), and '0' signifies that the subject was [censored](#). Censoring occurs when a subject leaves the study early, is still alive at the study's conclusion, or dies from a cause unrelated to the primary event. In these cases, we only know they survived up to the last observation time.

For our practical example, we will use a hypothetical clinical trial dataset. Column A represents the duration, measured in "Years in Trial," that each patient was followed. Column B records the event status, where '1' indicates the primary event (death) occurred, and '0' indicates the patient was censored (still alive or lost to follow-up). The initial setup of this data is crucial for the subsequent aggregation steps required by the Kaplan-Meier methodology.

	A	B
1	Years in trial	Died (1 = yes, 0 = no)
2	2	1
3	3	0
4	6	0
5	6	1
6	7	1
7	7	0
8	7	0
9	8	1
10	9	0
11	10	0
12	13	1
13	14	1
14	15	0
15	17	1
16	18	0

The core goal is to transform this raw, individual-level data into cumulative survival probabilities. This requires a transition from raw patient data to aggregated statistics based on specific time intervals where events were recorded. This transformation necessitates several intermediate calculation columns that adhere strictly to the statistical methodology underpinning the [Kaplan-Meier estimator](#).

## Formatting and Aggregating Survival Data

The [Kaplan-Meier method](#) calculates survival probabilities only at the specific time points where an event was observed. Therefore, the first critical step is to reorganize and condense the raw data into a timeline of unique observation points. We must extract all distinct "Years in Trial" values from Column A and list them sequentially in a new column (Column D). It is mandatory to include '0' as the starting time point, as the cumulative survival probability at time zero is definitionally 1.0 (or 100%).

**Step 1: List Unique Time Points.** Begin by listing all unique time values from the original raw data (Column A) into a new column, starting with '0'. These unique time points ( $D_i$ ) will serve as the reference points for all subsequent calculations. This aggregation drastically simplifies the dataset, focusing the analysis only on the moments when the risk profile changes.

	A	B	C	D
1	Years in trial	Died (1 = yes, 0 = no)		time
2	2	1		0
3	3	0		2
4	6	0		3
5	6	1		6
6	7	1		7
7	7	0		8
8	7	0		9
9	8	1		10
10	9	0		13
11	10	0		14
12	13	1		15
13	14	1		17
14	15	0		18
15	17	1		
16	18	0		

**Step 2: Calculate Key Survival Metrics.** Following the unique time points, we introduce four auxiliary columns (E, F, G, and H) to compute the necessary metrics: the number of events, the number of subjects at risk, the interval survival probability, and finally, the cumulative survival probability. Note that during this aggregation, it is efficient practice to exclude unique time points where no events occurred, as the survival probability remains constant during those periods. In the illustrated example, the time point "18" was removed from Column D because no corresponding death events were recorded at exactly 18 years.

### Calculating Key Survival Metrics using Excel Formulas

This phase utilizes powerful conditional counting and calculation functions in Excel to accurately replicate the [Kaplan-Meier estimator](#) methodology. Specifically, we rely on the [COUNTIFS](#) and [COUNTIF](#) functions to aggregate our raw data based on the unique time points established in Column D. A clear understanding of the definition of each resulting column is essential for proper interpretation.

	A	B	C	D	E	F	G	H
1	Years in trial	Died (1 = yes, 0 = no)		time	died	alive	1-(died/alive)	S(t)
2	2	1		0		15		1
3	3	0		2	1	15	0.9333333333	0.9333333333
4	6	0		3	0	14	1	0.9333333333
5	6	1		6	1	13	0.923076923	0.861538462
6	7	1		7	1	11	0.909090909	0.783216783
7	7	0		8	1	8	0.875	0.685314685
8	7	0		9	0	7	1	0.685314685
9	8	1		10	0	6	1	0.685314685
10	9	0		13	1	5	0.8	0.548251748
11	10	0		14	1	4	0.75	0.411188811
12	13	1		15	0	3	1	0.411188811
13	14	1		17	1	2	0.5	0.205594406
14	15	0						
15	17	1						
16	18	0						

The formulas below demonstrate the calculations required for the second time point (row 3) and the initial setup for the cumulative columns. Ensure that absolute references (using '\$') are correctly applied to the raw data range to prevent formula errors when dragging down.

**Column E (Number of Events, di):** This count represents the exact number of subjects who experienced the event (died) precisely at the time point specified in the corresponding row of Column D.

**E3 Formula:**

=COUNTIFS(\$A\$2:\$A\$16,D3,\$B\$2:\$B\$16,1)

**Column F (Number at Risk, ni):** This metric tallies the total number of subjects who were still under observation and potentially susceptible to the event just prior to the time point listed in Column D. The criteria ">"&D2-1 is used to count all observation times greater than or equal to the previous time point, effectively including all subjects still active.

**F2 Formula:**

=COUNTIF(\$A\$2:\$A\$16, ">"&D2-1)

**Column G (Interval Survival Probability, pi):** This calculates the conditional probability of surviving the current interval, given that the subject survived up to the beginning of this interval. It is derived by subtracting the hazard rate (events divided by those at risk) from 1.

**G3 Formula:** $=1-(E3/F3)$ 

**Column H (Cumulative Survival Probability,  $S(t)$ ):** This column is the core output of the Kaplan-Meier method, representing the overall probability of surviving beyond time  $t$ . It is calculated recursively as the product of all preceding interval survival probabilities.

**H2 Formula (Initial Value):** $=1$ **H3 Formula (Recursive Step):** $=H2*G3$ 

To complete the table, select the first calculated cell in each column (E3, F3, G3, H3) and drag the fill handle down to apply the formulas to the entire range of unique time points. Upon completion, Column H will contain the precise Y-axis values for the final survival curve visualization.

**Structuring Data for Visualization (Creating the Step Function)**

The resulting [survival curve](#) must visually appear as a step function: the survival probability remains flat between event times and drops vertically at the moment an event is observed. Standard Excel scatter plots connect points directly, which would result in a slanted line rather than the required step. To accurately generate this stepped appearance, we must duplicate and stagger our time (X) and survival (Y) values.

**Step 1: Copy Initial Results.** Create a new plotting area by copying the unique time points (Column D) into Column J, and the corresponding cumulative survival probabilities (Column H) into Column K. These two columns form the initial, un-staggered coordinate pairs.

D	E	F	G	H	I	J	K
time	died	alive	1-(died/alive)	S(t)		time	S(t)
0		15		1		0	1
2	1	15	0.9333333333	0.9333333333		2	0.933333
3	0	14	1	0.9333333333		3	0.933333
6	1	13	0.923076923	0.861538462		6	0.861538
7	1	11	0.909090909	0.783216783		7	0.783217
8	1	8	0.875	0.685314685		8	0.685315
9	0	7	1	0.685314685		9	0.685315
10	0	6	1	0.685314685		10	0.685315
13	1	5	0.8	0.548251748		13	0.548252
14	1	4	0.75	0.411188811		14	0.411189
15	0	3	1	0.411188811		15	0.411189
17	1	2	0.5	0.205594406		17	0.205594

**Step 2: Staggering the Data Points.** This is the most crucial manipulation for creating the step plot. The key is to insert a duplicate set of points that define the horizontal segment of the step. We achieve this by duplicating the time values (J3:J13 copied to J14:J24). However, we pair these duplicated time points with the survival probability values from the \*previous\* time interval (K2:K12 copied to K14:K24). This ensures that a coordinate pair exists to hold the line horizontally at  $S(t_{i-1})$  until time  $t_i$  is reached, at which point the line drops vertically.

J	K
time	S(t)
0	1
2	0.933333
3	0.933333
6	0.861538
7	0.783217
8	0.685315
9	0.685315
10	0.685315
13	0.548252
14	0.411189
15	0.411189
17	0.205594
2	1
3	0.933333
6	0.933333
7	0.861538
8	0.783217
9	0.685315
10	0.685315
13	0.685315
14	0.548252
15	0.411189
17	0.411189

**Step 3: Creating a Sort Index.** Because we duplicated and inserted data points non-sequentially, we need a mechanism to ensure Excel connects them in the correct chronological order (T0, T1, T1, T2, T2, etc.). Create a dedicated index column (Column L) that assigns a simple, increasing numerical value to each row in the new plotting range. This index is necessary to control the final sorting process.

J	K	L
time	S(t)	
0	1	1
2	0.933333	2
3	0.933333	3
6	0.861538	4
7	0.783217	5
8	0.685315	6
9	0.685315	7
10	0.685315	8
13	0.548252	9
14	0.411189	10
15	0.411189	11
17	0.205594	12
2	1	1
3	0.933333	2
6	0.933333	3
7	0.861538	4
8	0.783217	5
9	0.685315	6
10	0.685315	7
13	0.685315	8
14	0.548252	9
15	0.411189	10
17	0.411189	11

Create list of values in column L

Once the index is complete, highlight the entire plotting dataset (Columns J, K, and L) and apply a custom sort function. Sort the data based on the values in Column L, from smallest to largest. This step correctly interleaves the original and duplicated coordinate pairs, finalizing the data structure necessary for the step function plot.

J	K	L
time	S(t)	
0	1	1
2	1	1
2	0.933333	2
3	0.933333	2
3	0.933333	3
6	0.933333	3
6	0.861538	4
7	0.861538	4
7	0.783217	5
8	0.783217	5
8	0.685315	6
9	0.685315	6
9	0.685315	7
10	0.685315	7
10	0.685315	8
13	0.685315	8
13	0.548252	9
14	0.548252	9
14	0.411189	10
15	0.411189	10
15	0.411189	11
17	0.411189	11
17	0.205594	12

Sort columns J through L smallest to largest based on column L

## Generating the Final Survival Plot in Excel

With the X-coordinates (Time, Column J) and Y-coordinates (Cumulative Survival Probability, Column K) correctly ordered and staggered, we are ready to generate the final visual output of the [Kaplan-Meier estimator](#).

**Step 4: Inserting the Chart.** Highlight the finalized plotting data range (cells J2:K24). Navigate to the "Insert" tab in the Excel ribbon. Within the "Charts" group, select "Scatter" and specifically choose the "Scatter with Straight Lines and Markers" option. The straight lines ensure the points are connected in order, and the markers help identify the exact time points where the survival probability decreases, resulting in the desired step-function appearance.

The resulting chart provides a powerful visual summary of the subjects' survival experience. The curve always begins at 1.0 (representing 100% survival) at Time 0. Each downward step in the curve corresponds exactly to a time point where an event (death or failure) occurred, and the magnitude of the step is directly proportional to the hazard rate at that moment--the number of

events relative to the number of subjects at risk. Over time, the curve descends toward 0, reflecting the decreasing probability of survival.

To ensure the plot is professional and clear, customize the chart elements. It is vital to clearly label the axes: the X-axis should be labeled "Time" or "Years in Trial," and the Y-axis must be labeled "Cumulative Survival Probability." Adjusting line weight, color, and adding a descriptive chart title will further enhance the readability and impact of your completed survival curve, which should accurately resemble the following visualization:

