

Original Paper

Web-Based Survival Analysis Tool Tailored for Medical Research (KMplot): Development and Implementation

András Lánczky^{1,2}, MSc; Balázs Gyórfy^{1,2}, MD, PhD, DSC

¹Department of Bioinformatics, Semmelweis University, Budapest, Hungary

²TTK Lendület Cancer Biomarker Research Group, Institute of Enzymology, Research Centre for Natural Sciences, Budapest, Hungary

Corresponding Author:

Balázs Gyórfy, MD, PhD, DSC

Department of Bioinformatics

Semmelweis University

Tűzoltó utca 7-9

Budapest, 1094

Hungary

Phone: 36 305142822

Email: gyorffy.balazs@med.semmelweis-univ.hu

Abstract

Background: Survival analysis is a cornerstone of medical research, enabling the assessment of clinical outcomes for disease progression and treatment efficiency. Despite its central importance, no commonly used spreadsheet software can handle survival analysis and there is no web server available for its computation.

Objective: Here, we introduce a web-based tool capable of performing univariate and multivariate Cox proportional hazards survival analysis using data generated by genomic, transcriptomic, proteomic, or metabolomic studies.

Methods: We implemented different methods to establish cut-off values for the trichotomization or dichotomization of continuous data. The false discovery rate is computed to correct for multiple hypothesis testing. A multivariate analysis option enables comparing omics data with clinical variables.

Results: We established a registration-free web-based survival analysis tool capable of performing univariate and multivariate survival analysis using any custom-generated data.

Conclusions: This tool fills a gap and will be an invaluable contribution to basic medical and clinical research.

(*J Med Internet Res* 2021;23(7):e27633) doi: [10.2196/27633](https://doi.org/10.2196/27633)

KEYWORDS

Kaplan-Meier plot; internet; Cox regression; follow-up; multivariate analysis; survival

Introduction

Bioinformatic programs include databases, algorithms, services, and software tools. These not only span a wide range of utility but have also gained increased value in scientific research in recent years; approximately 80% of papers published in biology and 60% of papers published in medicine report the use of at least one bioinformatic tool [1]. We recently analyzed the landscape of web-based bioinformatic services and uncovered 3649 such publications since 1994, 69% of which are actively maintained [2]. The leading advantages of browser-based bioinformatic programs include unrestricted availability, the lack of need for the installation of specific software packages, optimized allocation of computational resources, the possibility of constant updates, instant access to the latest versions, and

the opportunity to enable real-time validation of previous analysis results. A subcohort of these tools enables certain analyses with user-provided data. A few representative examples used in these tools in medical research include an online calculator for receiver operator characteristics [3], a tool to determine optimal cut-off values for clinical tests [4], and a sample size calculator for randomized clinical trials [5].

The assessment of survival following the onset of a disease or of a treatment is a fundamental analysis in medical research. In an optimal scenario, the differential survival of two cohorts can be compared by employing a simple Mann-Whitney test. However, survival times do not follow a normal distribution and it is common for numerous subjects to lack associated event data at the end of follow-up. Kaplan and Meier [6] proposed a simple and elegant solution to these issues by including all cases

regardless of endpoint status in the analysis. The basic concept of Kaplan-Meier survival analysis is to assign a “censored” status to incomplete observations at the end of the follow-up time. In other words, there are two inputs for each case: the length of the follow-up time and a binary classifier designating the case as one with an event or one that is censored. Then, starting from 100, at each event, the survival line drops in proportion to the number of samples remaining within the investigated cohort. If there are multiple survival curves, the statistical difference between these is most commonly computed by employing the Cox proportional-hazards regression model [7].

Despite its widespread use, there is no online tool available for survival analysis. Therefore, it is necessary to acquire specialized software packages, as none of the general office packages (OpenOffice, LibreOffice, MS Office) is suitable for analyzing follow-up data. We previously established an online platform capable of linking survival outcome in various cancer types to mRNA [8] and microRNA [9] expression alterations. Here, we aimed to establish a freely available, easy-to-use online platform capable of performing survival analysis and constructing a Kaplan-Meier plot with any type of user-uploaded custom data containing any type of genomic or clinical information.

Methods

Setup of the Web Platform

The website is built on an Apache 2.4 web server and hosted by a Linux-based server machine. The user interface is written

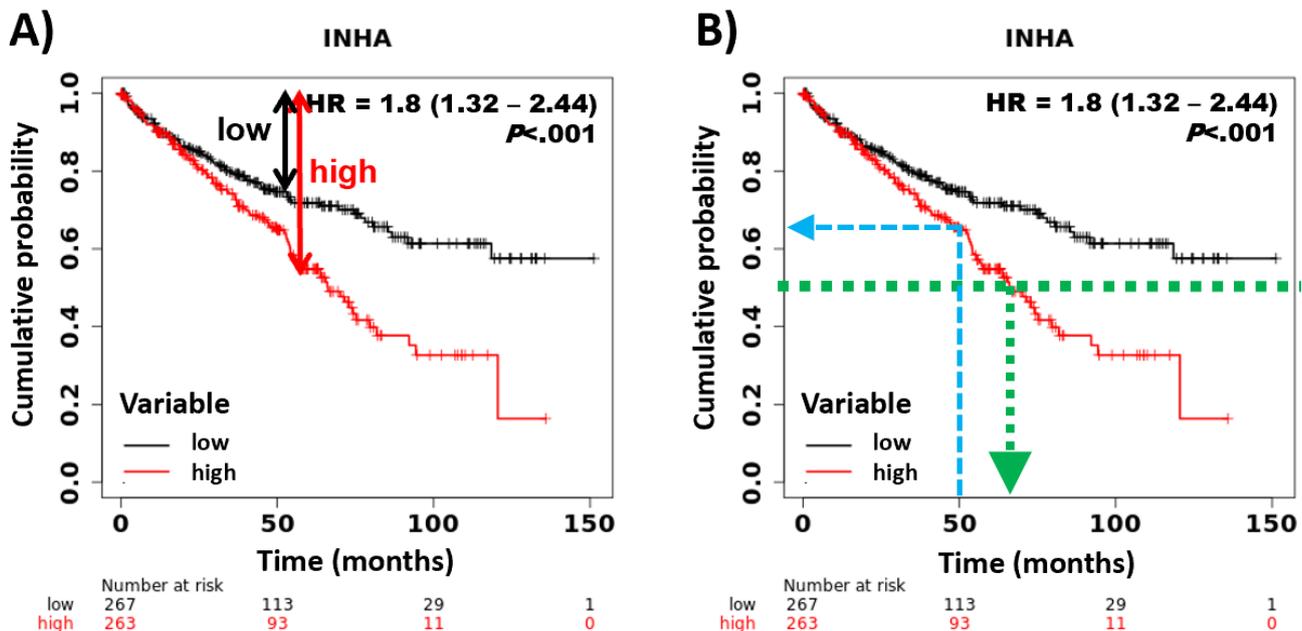
in PHP 7 and JavaScript using JQuery. The backend side is written in PHP 7 and R, and the repository layer is built on the PostgreSQL 12 database. The database temporarily contains the uploaded data and generated results. The analysis platform is accessible via any standard browser (Firefox, Edge, Chrome, Safari).

Survival Analysis

Multiple R packages are used for the statistical computations and for generating the output graphs. The *survival* package [10] is used for univariate Kaplan-Meier analysis and the multivariate analysis. The survival curve and the beeswarm plot are generated by the *survplot* [11] and *beeswarm* [12] packages, respectively. The *XML* and *rjson* R packages are used to load the configuration files, the *RODBC* package is used to communicate with the database, and the *ggplot2* package [13] is used to visualize the results.

When comparing two cohorts, the significance is computed using the Cox-Mantel (log rank) test [7]. The difference between the cohorts is numerically characterized by the hazard rate (HR), which is based on the differential descent rate of the two cohorts (Figure 1A). Of note, since the hazard rate is by definition a comparison to the baseline, a relative two-fold drop in one cohort is equal to a half-fold drop in the other cohorts. Basically, depending on the context, an HR of 2 equals an HR of 0.5. As it is easier to understand an HR value above 1 in most cases, we implemented an option to invert all HR values below 1.

Figure 1. Kaplan-Meier curves showing main concepts used in survival analysis, including the (A) hazard rate (high/low) and (B) median survival. The green arrow shows the visually determined median survival and the blue arrow shows the survival probability at 50 months.



The generated results also include the median survival time, which is the time at which the probability of 0.5 is reached in one of the cohorts. The median time can also be determined visually by drawing a vertical line from the selected probability

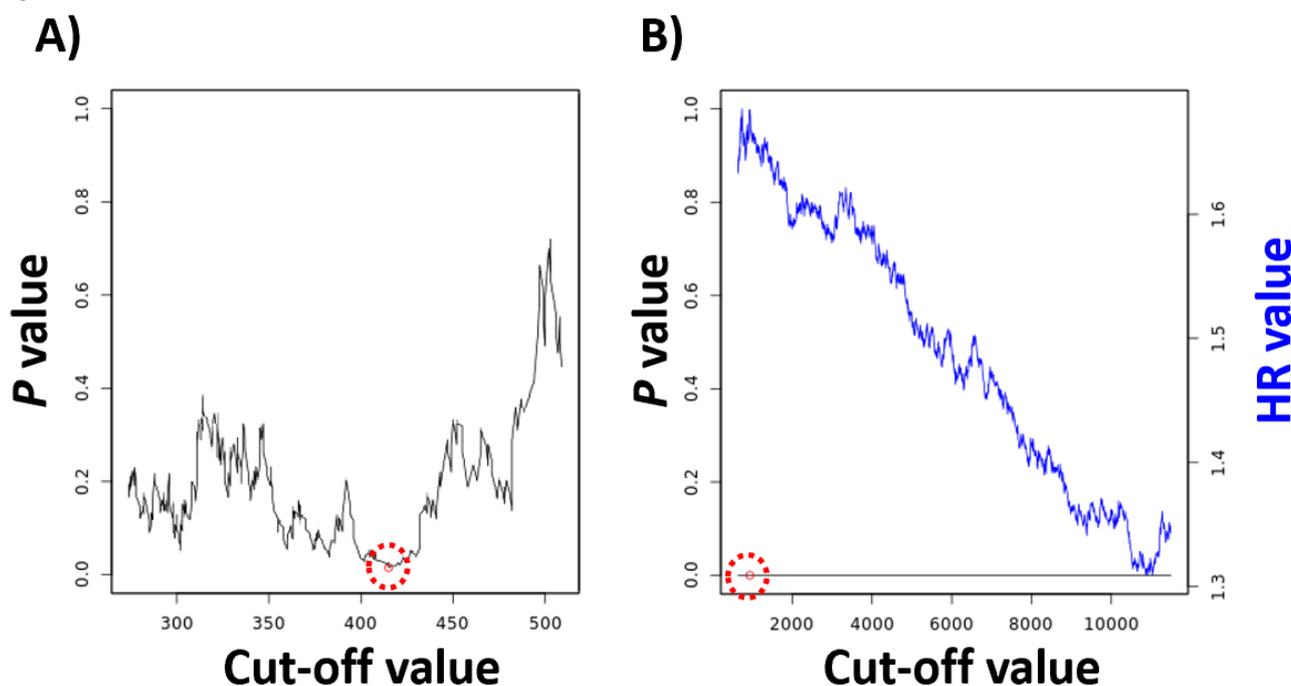
to the X axis. Of note, performing the steps backward can determine the cumulative probability of survival at a given time point (Figure 1B).

Assigning the Samples into Two Cohorts

To enable visualization in the Kaplan-Meier plot, it is necessary to establish a cut-off value and assign the samples to one of two cohorts. We implemented three different options for this task: using a predefined quantile (including the median, upper, and lower quartiles), trichotomizing the data (eg, assign the data into three cohorts and then omit the middle cohort), and using the best available cut-off value.

To find the best cutoff, we iterate over the input variable values from the lower quartile to the upper quartile and compute the Cox regression [7] for each setting. The most significant cut-off value is used as the best cutoff to separate the input data into two groups. We implemented a simple visual representation of this analysis by showing the achieved P values in relation to the used cut-off values (Figure 2A). In case the generated cut-off values are ambiguous (eg, multiple cut-off values deliver very low P values), the cut-off value corresponding to the highest HR is used (see Figure 2B).

Figure 2. A cut-off plot can be used to visualize the correlation between the used cut-off values and the achieved P values (A) and hazard rate (HR) (B). The red circle identifies the best cutoff. The computation of false discovery rate across all P values provides correction for multiple hypothesis testing.



Quality Control

During the computation of multiple cut-off values, multiple hypotheses are generated. Therefore, the false discovery rate (FDR) is computed by default in this setting using the Benjamini-Hochberg method [14] to correct for multiple hypothesis testing. The FDR results are normally shown in the "Results" page.

A requirement for Cox regression is that the hazard is independent of time. To fulfill this requirement, the censoring should be independent of the prognosis, samples entering at different time points in the analysis should have the same prognosis, and the time should be measured as a continuous variable (not in bins). We employed the *coxph* function of the *survival* package [10] for performing the proportional hazard assumption test.

In some cases, one might want to compare clinical and genomic variables. To enable this, clinical data can be selected not only as filters but also as variables to be included in the multivariate analysis. In these analyses, the "Results" page displays the P values and HR values for each variable included in the multivariate analysis in a table format.

Using Multiple Variables

We implemented multiple options to simultaneously use and combine multiple variables. Each of these settings uses the original variable values as input and basic mathematical functions to calculate the new joint values.

The simplest option is to select multiple variables and then use each variable separately. In this case, the same analysis is performed for each selected marker using the exact same filtering settings. This option is identical to running the analysis for each variable consecutively.

In the second feature, one can use the mean expression of a panel of variables; in this case, any variable can be inverted and a weight can be added to each. Using the mean expression of a set of genes can be termed a "signature analysis," as the expression of each included variable will influence the value of the final "composite variable." This feature can also be used to validate previously published gene expression signatures utilizing a preselected panel of genes.

A third option is utilization of the ratio of two genes; in this case, one variable is used as the numerator, the other variable is used as the denominator, and a new value is computed for

each sample. This setting is useful when one wants to compare the expression values to a reference gene such as *GAPDH*.

The fourth option enables the stratification of all patients based on the median expression level of a selected variable and then use another variable in the high or low cohort only. This enables the investigation of a selected variable in an already stratified cohort and ultimately the setup of a decision tree-like classification for the investigated cohort.

In each of the settings where multiple variables are combined, a new value based on the equation is generated for each sample, which is then used when performing the survival analysis, including the cut-off selection. Of note, one might want to directly compare two or more selected continuous variables to each other. For this purpose, we implemented an option to compute Spearman and Pearson correlation coefficients between the variables using the *cor.test* function from the basic R distribution.

Results

We established an online survival analysis platform that grabs a user-generated tab-separated or semicolon-separated file as input. The table headers can include case-insensitive letters of the English alphabet, numbers, spaces, underscores, colons,

round brackets, and exclamation marks as characters. The content within the table cells can be numeric or text values. Some columns can be used as filters and a maximum of three filters are allowed. Table 1 provides a quick guide for the setup of an input file. The file can be a comma-separated value or a tab-separated table, and different types of data are allowed in each column. Table 2 shows a sample input file using this guide; the maximal dimensions of the table are 100 columns and 8000 rows. Note that a gene can be in the form of text when used as a group. Using this data table, the system is capable of performing univariate and multivariate survival analysis by using one or multiple variables and clinical data. In addition to drawing a Kaplan-Meier plot, the *P* values and HR values with 95% CIs are also computed. A separate plot visualizes the correlation between the *P* values and HR values and the employed cut-off values. Median survival values are computed for cohorts reaching a cumulative probability of 0.5, and upper-quartile survival is computed for the remaining cases. Of note, when performing multivariate analysis, only patient samples for which all variables of interest are concurrently available can be included. The platform includes multiple quality-control steps, including validation of the proportional hazard assumption and computation of the FDR for cases where multiple analyses are run simultaneously. The web service is freely available without requiring registration [15].

Table 1. Quick start guide for setting up an input file.

Header name	Sample ID	Survival time	Survival event	Filter	Gene
Automatically recognized	Yes	Yes	Yes	Yes	No
Maximal number of different values	No limit	No limit	2 (0 or 1)	10	No limit
Can be text	Yes	No	No	Yes	No
Can be binary	No	No	Yes	Yes	Yes
Can be continuous		Yes	No	No	Yes

Table 2. Sample input file.

Sample ID	Survival time	Survival event	Filter_A	Filter_B	Filter_C	Gene_1	ABC123	DE45
Sample 1	95	1	2	2	3	1441	4474	1.13
Sample 2	66	0	3	3	3	3064	421	2.395
Sample 3	70	0	3	1	1	2529	2974	1.363
Sample 4	26	1	3	1	3	19	3346	4.818
Sample 5	13	0	1	2	3	3573	1244	2.058
Sample 6	67	0	2	3	2	2977	962	4.431
Sample 7	96	1	3	3	3	2777	4367	2.015
Sample 8	67	0	3	3	1	4606	4190	1.05
Sample 9	95	1	3	1	2	1209	3930	1.980
Sample 10	1	1	2	3	2	1894	4897	4.073

Discussion

Currently, genomics, transcriptomics, proteomics, and metabolomics enable the simultaneous investigation of multiple markers related to patient prognosis in experimental and clinical

studies. Multiple online tools make survival analysis possible using previously published datasets such as those employing data from The Cancer Genome Atlas [9,16]. Despite the almost ubiquitous use of Cox regression to correlate different marker levels to prognosis, there is no available software to perform survival analysis for user-generated custom datasets. We

established a wide-ranging online tool capable of performing Cox regression and constructing Kaplan-Meier plots for user-generated data. A comprehensive and practical review of the Kaplan-Meier curves has been published previously [17].

A major advantage of our platform is the inclusion of multiple choices to select a cut-off value to be used in the analysis. To generate a Kaplan-Meier plot, one must first determine a cutoff; a convenient and widespread option for this task is the median expression value [18,19]. However, the cutoff should be based on the intention of the study. In most medical studies, there is no biological reason that a certain predetermined quantile cutoff should discriminate two cohorts [20]. When a researcher aims to uncover any potential correlation between a variable and outcome, then all possible cut-off values can be checked. Of course, in such cases, the chance of false-positive results also increases; therefore, we have implemented the Benjamini-Hochberg method [14] to calculate the FDR to correct for multiple hypothesis testing. Our approach is rather conservative as the different analyses are not truly independent in such a scenario, as only a few samples can switch cohorts between successive analyses. Of note, independent of the used cutoff, a single-variable analysis is almost never sufficient to prove a direct correlation and thus multivariate analysis should not be omitted.

The analysis automatically checks the proportional hazards assumption to evaluate the independence from time. This can also be achieved by a simple visual inspection of the graph: in case there seems to be a significant difference between the two cohorts but the lines cross at multiple time points, then the hazard is clearly not independent of time [21]. Of note, a common question is whether or not crossing at the right end of the plot violates the proportional hazards assumption. In most cases, at the end of the follow-up time, only few patients remain in both cohorts. Thus, because the drop in the line for each event is proportional to all samples remaining in the analysis, even an event for a single patient can result in crossing of the two lines. However, this will not affect the significance of the entire analysis.

When interpreting the results, one has to be aware of some common caveats of survival analysis. First, the *P* value should be interpreted with respect to the sample size. The Cox model is not suitable for small sample sizes ($N < 40$), and in these cases the generalized log-rank method is a better choice [22]. Higher sample numbers will lead to better significance, even in cases where the HR values are lower. A representative example of

this bias is the ill-fated FLEX phase III trial [23]. By investigating the effect of cetuximab in patients with advanced nonsmall cell lung cancer, the authors observed a difference in survival of 10.1 months vs 11.3 months in the untreated and treated cohort, respectively. Although this difference was initially considered to be sufficient to gain approval by the US Food and Drug Administration, the European Medicines Agency rejected approval of the drug. Their main problem with the trial was the minimal overall survival benefit of only 12% and that only the exceptionally high sample number ($N=1125$) helped to reach a minimally significant *P* value of .04 [23].

A second important deception is the proportion of recorded events within a study. As only the actual events contribute to the drops in survival curves, it is not possible to perform a meaningful survival analysis when the number of events is very low. This not only prevents the computation of median (or upper quartile) survival, but the accidental concentration of all events into one of the cohorts can lead to an infinite HR. For example, The Cancer Genome Atlas Network published a breast cancer dataset with approximately 1000 patient samples [24]. The authors had to note that because of the very short follow-up, only 11% of the samples had survival events, which prevented utilization of the dataset for survival analyses [24].

We also have to discuss some limitations of the software. The input file has to be carefully formatted, and a maximum of 100 columns and 8000 rows are allowed. Only full columns are acceptable as variables, a maximum of three filters can be defined, and the survival event can only be coded "0" or "1." Although these restrictions can make the setup of the analysis challenging, when a correctly formatted table is uploaded, the system can automatically recognize columns representing a survival event or survival time. A second limitation is the exclusive use of the Cox proportional-hazards model to compute significance, and other tests such as the Cochran-Mantel-Haenszel test [25,26] or the Gehan-Breslow-Wilcoxon test [27,28] are not implemented. The reason for our restriction is the almost exclusive use of the Cox test in the current medical literature.

In summary, we established an online survival analysis tool capable of performing univariate and multivariate survival analysis using any custom-generated data. We believe that this registration-free online platform simultaneously integrating multiple different analysis and quality-control options will be a valuable tool for biomedical researchers.

Acknowledgments

The research was financed by the 2018-2.1.17-TET-KR-00001, 2020-1.1.6-JÖVŐ-2021-00013 and 2018-1.3.1-VKE-2018-00032 grants and by the Higher Education Institutional Excellence Program (2020-4.1.1.-TKP2020) of the Ministry for Innovation and Technology (MIT) in Hungary, within the framework of the Bionic thematic program of the Semmelweis University. The authors wish to acknowledge the support of ELIXIR Hungary.

Conflicts of Interest

None declared.

References

1. Duck G, Nenadic G, Filannino M, Brass A, Robertson DL, Stevens R. A survey of bioinformatics database and software usage through mining the literature. *PLoS One* 2016;11(6):e0157989 [FREE Full text] [doi: [10.1371/journal.pone.0157989](https://doi.org/10.1371/journal.pone.0157989)] [Medline: [27331905](https://pubmed.ncbi.nlm.nih.gov/27331905/)]
2. Ósz Á, Pongor LS, Szirmai D, Gyórfy B. A snapshot of 3649 web-based services published between 1994 and 2017 shows a decrease in availability after 2 years. *Brief Bioinform* 2019 May 21;20(3):1004-1010 [FREE Full text] [doi: [10.1093/bib/bbx159](https://doi.org/10.1093/bib/bbx159)] [Medline: [29228189](https://pubmed.ncbi.nlm.nih.gov/29228189/)]
3. Goksuluk D, Korkmaz S, Zararsiz G, Karaagaoglu A. easyROC: an interactive web-tool for ROC curve analysis using R language environment. *R Journal* 2016;8(2):213. [doi: [10.32614/rj-2016-042](https://doi.org/10.32614/rj-2016-042)]
4. Budczies J, Klauschen F, Sinn BV, Gyórfy B, Schmitt WD, Darb-Esfahani S, et al. Cutoff Finder: a comprehensive and straightforward Web application enabling rapid biomarker cutoff optimization. *PLoS One* 2012;7(12):e51862 [FREE Full text] [doi: [10.1371/journal.pone.0051862](https://doi.org/10.1371/journal.pone.0051862)] [Medline: [23251644](https://pubmed.ncbi.nlm.nih.gov/23251644/)]
5. Wang X, Ji X. Sample size estimation in clinical research: from randomized controlled trials to observational studies. *Chest* 2020 Jul;158(1S):S12-S20. [doi: [10.1016/j.chest.2020.03.010](https://doi.org/10.1016/j.chest.2020.03.010)] [Medline: [32658647](https://pubmed.ncbi.nlm.nih.gov/32658647/)]
6. Kaplan E, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958 Jun;53(282):457-481. [doi: [10.2307/2281868](https://doi.org/10.2307/2281868)]
7. Cox DR. Regression models and life-tables. *J R Stat Soc B* 2018 Dec 05;34(2):187-202. [doi: [10.1111/j.2517-6161.1972.tb00899.x](https://doi.org/10.1111/j.2517-6161.1972.tb00899.x)]
8. Gyórfy B, Lánczky A, Szállási Z. Implementing an online tool for genome-wide validation of survival-associated biomarkers in ovarian-cancer using microarray data from 1287 patients. *Endocr Relat Cancer* 2012 Apr;19(2):197-208. [doi: [10.1530/ERC-11-0329](https://doi.org/10.1530/ERC-11-0329)] [Medline: [22277193](https://pubmed.ncbi.nlm.nih.gov/22277193/)]
9. Lánczky A, Nagy Á, Bottai G, Munkácsy G, Szabó A, Santarpia L, et al. miRpower: a web-tool to validate survival-associated miRNAs utilizing expression data from 2178 breast cancer patients. *Breast Cancer Res Treat* 2016 Dec;160(3):439-446. [doi: [10.1007/s10549-016-4013-7](https://doi.org/10.1007/s10549-016-4013-7)] [Medline: [27744485](https://pubmed.ncbi.nlm.nih.gov/27744485/)]
10. Therneau TM, Grambsch PM. Modeling survival data: extending the Cox model. New York, NY: Springer-Verlag; 2000.
11. Survplot. URL: <http://www.cbs.dtu.dk/~eklund/survplot> [accessed 2021-07-07]
12. BeeSwarm. URL: <http://www.cbs.dtu.dk/~eklund/beeswarm/> [accessed 2021-07-07]
13. Wickham H. ggplot2: elegant graphics for data analysis. New York, NY: Springer-Verlag; 2009.
14. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc B* 2018 Dec 05;57(1):289-300. [doi: [10.1111/j.2517-6161.1995.tb02031.x](https://doi.org/10.1111/j.2517-6161.1995.tb02031.x)]
15. Kaplan-Meier plotter with custom data. URL: https://kmplot.com/custom_data [accessed 2021-07-02]
16. Pak K, Oh S, Goh T, Heo H, Han M, Jeong D, et al. A user-friendly, web-based integrative tool (ESurv) for survival analysis: development and validation study. *J Med Internet Res* 2020 May 05;22(5):e16084 [FREE Full text] [doi: [10.2196/16084](https://doi.org/10.2196/16084)] [Medline: [32369034](https://pubmed.ncbi.nlm.nih.gov/32369034/)]
17. Rich JT, Neely JG, Paniello RC, Voelker CCJ, Nussenbaum B, Wang EW. A practical guide to understanding Kaplan-Meier curves. *Otolaryngol Head Neck Surg* 2010 Sep;143(3):331-336 [FREE Full text] [doi: [10.1016/j.otohns.2010.05.007](https://doi.org/10.1016/j.otohns.2010.05.007)] [Medline: [20723767](https://pubmed.ncbi.nlm.nih.gov/20723767/)]
18. Mishra NK, Southekal S, Guda C. Survival analysis of multi-omics data identifies potential prognostic markers of pancreatic ductal adenocarcinoma. *Front Genet* 2019 Jul 18;10:624. [doi: [10.3389/fgene.2019.00624](https://doi.org/10.3389/fgene.2019.00624)] [Medline: [31379917](https://pubmed.ncbi.nlm.nih.gov/31379917/)]
19. Liu K, Jin M, Xiao L, Liu H, Wei S. Distinct prognostic values of mRNA expression of glutathione peroxidases in non-small cell lung cancer. *Cancer Manag Res* 2018 Aug;10:2997-3005. [doi: [10.2147/cmar.s163432](https://doi.org/10.2147/cmar.s163432)]
20. Giannoni A, Baruah R, Leong T, Rehman MB, Pastormerlo LE, Harrell FE, et al. Do optimal prognostic thresholds in continuous physiological variables really exist? Analysis of origin of apparent thresholds, with systematic review for peak oxygen consumption, ejection fraction and BNP. *PLoS One* 2014;9(1):e81699 [FREE Full text] [doi: [10.1371/journal.pone.0081699](https://doi.org/10.1371/journal.pone.0081699)] [Medline: [24475020](https://pubmed.ncbi.nlm.nih.gov/24475020/)]
21. Hess K. Graphical methods for assessing violations of the proportional hazards assumption in Cox regression. *Stat Med* 1995 Aug 15;14(15):1707-1723. [doi: [10.1002/sim.4780141510](https://doi.org/10.1002/sim.4780141510)] [Medline: [7481205](https://pubmed.ncbi.nlm.nih.gov/7481205/)]
22. Xu R, Shaw PA, Mehrotra DV. Hazard ratio estimation in small samples. *Stat Biopharm Res* 2018 Jun 20;10(2):139-149. [doi: [10.1080/19466315.2017.1369899](https://doi.org/10.1080/19466315.2017.1369899)]
23. Pirker R, Pereira JR, Szczesna A, von Pawel J, Krzakowski M, Ramlau R, FLEX Study Team. Cetuximab plus chemotherapy in patients with advanced non-small-cell lung cancer (FLEX): an open-label randomised phase III trial. *Lancet* 2009 May 02;373(9674):1525-1531. [doi: [10.1016/S0140-6736\(09\)60569-9](https://doi.org/10.1016/S0140-6736(09)60569-9)] [Medline: [19410716](https://pubmed.ncbi.nlm.nih.gov/19410716/)]
24. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* 2012 Oct 04;490(7418):61-70 [FREE Full text] [doi: [10.1038/nature11412](https://doi.org/10.1038/nature11412)] [Medline: [23000897](https://pubmed.ncbi.nlm.nih.gov/23000897/)]
25. Cochran WG. Some methods for strengthening the common χ^2 tests. *Biometrics* 1954 Dec;10(4):417. [doi: [10.2307/3001616](https://doi.org/10.2307/3001616)]
26. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst* 1959 Apr;22(4):719-748. [doi: [10.1093/jnci/22.4.719](https://doi.org/10.1093/jnci/22.4.719)] [Medline: [13655060](https://pubmed.ncbi.nlm.nih.gov/13655060/)]
27. Gehan EA. A generalized two-sample Wilcoxon test for doubly censored data. *Biometrika* 1965 Dec 01;52(3-4):650-653. [doi: [10.1093/biomet/52.3-4.650](https://doi.org/10.1093/biomet/52.3-4.650)]

28. Breslow N. A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship. *Biometrika* 1970;57(3):579-594. [doi: [10.1093/biomet/57.3.579](https://doi.org/10.1093/biomet/57.3.579)]

Abbreviations

FDR: false discovery rate

HR: hazard rate

Edited by R Kukafka; submitted 01.02.21; peer-reviewed by T Karn, YH Kim; comments to author 15.02.21; revised version received 19.02.21; accepted 06.05.21; published 26.07.21

Please cite as:

Lánczky A, Gyórfy B

Web-Based Survival Analysis Tool Tailored for Medical Research (KMplot): Development and Implementation

J Med Internet Res 2021;23(7):e27633

URL: <https://www.jmir.org/2021/7/e27633>

doi: [10.2196/27633](https://doi.org/10.2196/27633)

PMID: [34309564](https://pubmed.ncbi.nlm.nih.gov/34309564/)

©András Lánczky, Balázs Gyórfy. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 26.07.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.