

# “Crying like a Baby”: Survival Analysis and the Multimodal Memory of Holocaust Survivors

Gabor Mihaly Toth<sup>1</sup> , Mohamed Laib<sup>2</sup> , Alina Bothe<sup>3</sup> , Marcus Ma<sup>4</sup> , Shrikanth Narayanan<sup>4</sup> , Cedric Pruski<sup>2</sup> , and Marcos da Silveira<sup>2</sup> 

<sup>1</sup> Center for Contemporary and Digital History, University of Luxembourg, Luxembourg

<sup>2</sup> Luxembourg Institute of Science and Technology, Luxembourg

<sup>3</sup> Selma Stern Center for Jewish Studies Berlin-Brandenburg, Free University, Berlin, Germany

<sup>4</sup> Viterbi School of Engineering, University of Southern California, Los Angeles, United States

## Abstract

This paper presents a novel application of survival analysis to study gender differences in emotional responses during Holocaust survivor testimonies. By analyzing 982 oral history interviews (454 women, 528 men) from the USC Shoah Foundation, we examine the temporal patterns of crying and sobbing events using multimodal analysis techniques. Our findings reveal that women survivors cry earlier and more frequently than men during their testimonies, with statistically significant differences in waiting times and recurrence patterns. Furthermore, we identify gender-specific triggers: women are more likely to cry when discussing forced labor, while men show emotional responses when discussing captivity, mistreatment, and religious topics. This study demonstrates how survival analysis can provide meaningful insights into the temporality and gendered nature of traumatic memory transmission in oral history interviews.

**Keywords:** Holocaust testimonies, survival analysis, multimodal analysis, gender differences, oral history, emotional response

## 1 Introduction

It is the summer of 1944. Freight trains packed with men, women, and children arrive at Auschwitz-Birkenau. Families are forcibly separated and sent to their death. Those few who survived have recounted this moment millions of times. So did Alex and Elisabeth. They both arrived in Birkenau around the same time - though from different places. Fifty years later, they both gave oral history testimonies. However, the recollection by Elisabeth is different from the way Alex recalls the moment of forced separation. Elisabeth breaks down. She keeps crying when recalling the moment she was separated from her parents and her brother. They did not survive. Alex recalls the moments of separation from his parents without any visible emotional breakdown. Neither did his parents survive. Genuinely, one asks the question: are the different recollections by Elisabeth and Alex part of a broader trend or pattern? In their belated oral history testimonies, do female and male survivors of Nazi concentration camps start to cry and break down at the recollection of different episodes? Are women or men more likely to cry when addressing a troubling past? Throughout the interview process, do men or women tend to be overwhelmed earlier? The relevance of these questions is underscored by a leitmotif, a phrase that Holocaust survivors keep repeating in their testimonies: “crying like a baby.”

---

Gabor Mihaly Toth, Mohamed Laib, Alina Bothe, Marcus Ma, Shrikanth Narayanan, Cedric Pruski, and Marcos da Silveira. ““Crying like a Baby”: Survival Analysis and the Multimodal Memory of Holocaust Survivors.” In: *Computational Humanities Research 2025*, ed. by Taylor Arnold, Margherita Fantoli, and Ruben Ros. Vol. 3. Anthology of Computers and the Humanities. 2025, 4–15. <https://doi.org/10.63744/sTRxoCHdBgK5>.

© 2025 by the authors. Licensed under Creative Commons Attribution 4.0 International (CC BY 4.0).

Holocaust scholars have been studying both the verbal and the non-verbal aspects of survivors' memory as transmitted in belated audio/video testimonies; they have been investigating the performative aspects of recounting the past [23]. A multimodal approach focusing on the fusion of and interrelations between different modalities through which traumatic narratives unfold has been applied since the early 1990s [10; 17]. Nonetheless, previous studies worked with a limited number of testimonies and approached multimodality through the close-reading and close-listening of single testimonies [12; 20]. A large-scale multimodal analysis of oral history interviews with Holocaust survivors was not possible for two key reasons. First, even though approximately 100.000 testimonies have been recorded since the end of the Second World War, these testimonies were not available as raw data; researchers could interact with them only through online databases, which do not allow large-scale data analysis. Second, neither transcripts nor multimodal annotation, highlighting for instance moments of crying, were available. All this is now changing. Testimonies are increasingly more available as data sets and generative AI facilitates their multimodal annotation on an unprecedented scale [7]. Interest in multimodality is gaining momentum in the Humanities and beyond [26; 27]. Despite this recent development, there is still a key challenge to be tackled before large-scale multimodal studies of Holocaust memory can take place.

The challenge that Holocaust scholars, as well as students of other oral history interviews face, is the lack of a comprehensive statistical framework for analyzing the multimodality of audio/video interviews. A key goal of the VOICES project (<https://voices.uni.lu>), which the authors of this paper are affiliated with, is to develop components for this framework. Here, we present one of our components that relies on different branches of survival analysis. This is a statistical tool widely applied in clinical research and engineering [15]. By analyzing 982 Holocaust testimonies (454 women, 528 men) (timestamped transcripts with multimodal clues inserted, as well as topical annotation added, see the data description in the Appendix) we address questions related to gender and crying. Using survival analysis and other statistical tools, we find both unimodal and multimodal evidence that women survivors cried more often and earlier than men.

## 2 The problem set and survival analysis

An oral history interview as a time-continuous multimodal data can be conceptualized as parallel sequences of events. Each sequence is a layer focusing on different features of the original audio/video data. For instance, the rate of speech is one layer; the movement of the speakers' eye forms another layer. Change of posture or the abrupt acceleration of speech rate are distinct events of the respective sequences. In addition to the non-verbal layers, extrapolated from voice, posture, and other physical and biological aspects of the speakers, another key - verbal - layer is the content of what is discussed. This layer can be also viewed as a sequence of events; each topic discussed is an event belonging to the verbal layer. In an oral history interview, the different verbal and non-verbal layers are parallel to each other since they unfold along the same temporal axis. If multiple speakers (for instance an interviewer and several interviewees) are involved, each speaker can be seen as a different set of layers.

It is reasonable to assume that different layers of multimodality are interconnected with each other. For instance, certain events, that is to say topics, on the content layer trigger specific events on the non-verbal layers. Explained through the example of the two Auschwitz survivors, the narration of forceful separation upon arrival to the death camp triggers a specific vocal event, which is crying. In Section 5, we present the results of our analysis investigating the interconnections between events belonging to two different layers of multimodality by one speaker. At the same time, here we do not address the problem of how different sets of layers by different speakers are related to each other; this is subject of further research.

It is important to note that the occurrence of an event belonging to a given single layer of multimodality involves a specific type of time. This is waiting time. For instance, as we note in

Section 3, survivors do not start crying at the beginning of an interview. One has to listen and keep waiting for this event to occur. The “usual” time one needs to wait for an event is informative. For instance, it gives access to the way interviews by different groups of people, such as for instance men and women, tend to unfold. As we demonstrate in the following sections, by focusing on the waiting time, we can address the research questions raised in the introduction. For instance, we can study whether men or women are more likely to break down emotionally throughout the interview process. For studying the waiting time for the event of crying, we rely on survival analysis.

Specifically, with survival analysis we can address the following broader but more technical questions that fall into two groups. The first group focuses on the study of a single multimodal layer, such as crying and not crying, and works with the waiting time for the first occurrence of an event: how long is the waiting time for a woman or a man to start to cry after an interview begins? The second group also works with a single layer but focuses on the waiting time until multiple events of the same type occur after the beginning of an interview. With this, we can study the frequency of crying and its patterns of recurrence in interviews by women and men. With the focus on the waiting time, survival analysis brings a great variety of quantitative and qualitative phenomena into play, which enrich our understanding of multimodal memory transmitted through oral history interviews.

Finally, survival analysis helps address a key problem. This is the perceived randomness of the waiting time for events to occur, which can be understood through its application in clinical trials. For instance, there is a clinical trial investigating the effectiveness of a new method to treat terminal cancer; the goal of the trial is to find out whether the treatment is able to prolong the life of those one thousand men and women who participated. The trial aims to give explicit information about the time one is expected to live after the treatment; furthermore, it wants to compare the effectiveness of treatment for men and women. Nonetheless, the time until death varies a lot among the participants. From the perspective of the researchers, the duration (or the waiting time) seems to be a random phenomenon that changes from subject to subject. Survival analysis models the waiting time for one or multiple events to occur as a random variable; it uses the data collected throughout the clinical trial to infer information about the waiting time as a random variable, including the underlying probabilistic processes. Interestingly, the usual goal in a clinical trial, modelling the perceived randomness of the waiting time until one or multiple events occur, has similarities with the occurrence of multimodal events in oral history interviews. Each interview can be conceptualized as one subject in a clinical trial. If the focus is the single layer (or unimodality) defined by for example the events of crying and not\_crying, the goal of the researcher is to model the perceived randomness of the usual waiting time - calculated from the beginning of the interview - for an interviewee to cry for the first time or for a number of times. The first case (focus on the first event) belongs to survival analysis; the second case (focus on multiple events) belongs to recurrent events analysis, which is a specific branch of survival analysis [2]. In case the focus is on the relationship between two layers, the waiting time as a random variable is to be measured from a starting point on one layer until response belonging to the other layer. As a whole, both clinical trials and oral history interviews involve waiting time as a random variable; survival analysis enables us to model this random variable, and describe its qualitative and quantitative aspects. It is an established framework equipped with statistical significance testing and group comparison, which, as we will see below, are highly useful to compare and gather insights into multimodal and gendered memory of the Holocaust transmitted through oral history interviews with survivors.

In the next sections, we will discuss how survival analysis supports the study of  
A, a single layer of multimodality and the waiting time for the first event defined by crying;  
B, a single layer of multimodality and the waiting time for the multiple crying events to occur  
To study the interrelationship between events (crying and topics) on two layers of multimodal-

ity, we use another statistical tool, odds ratio analysis, explained in the Methods section. First, we present the results of our analysis. In the Methods section, we give a technical description. Finally, in the Discussions, we address the strengths and the limitations of the survival analysis for the use of multimodal data analysis, and discuss future research directions.

### 3 Results: single layer and the first crying and sobbing event

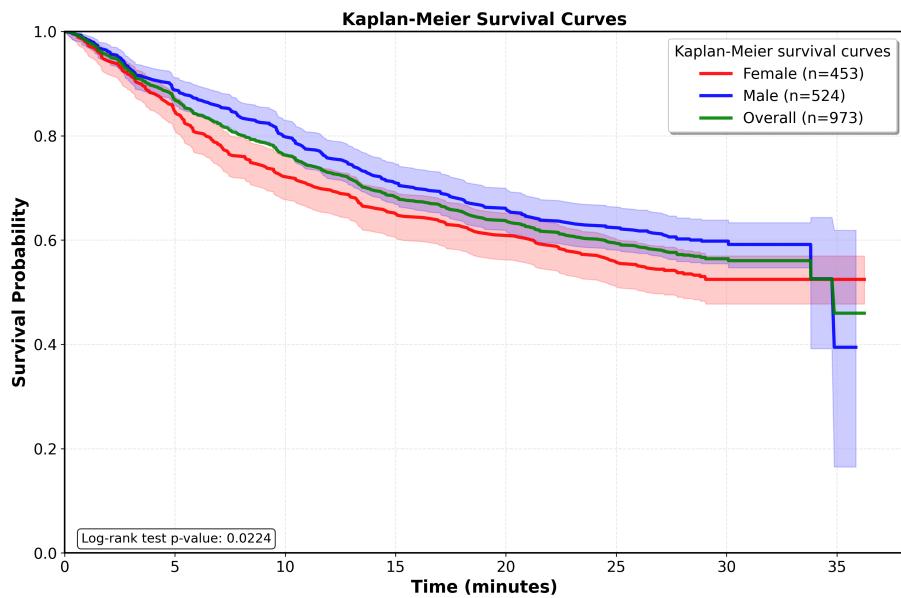
Even though it is a natural phenomenon that survivors of Nazi concentration camps cry when they recall a traumatic past, it is not always the case. In our interview data set, less than half of the participants actually cry or sob - corresponding to 43.8% event rate. Given that there is no statistically significant difference between the event rate of crying and sobbing separately for men and women (results show p-value of 0.14 for crying event only). Therefore, we combined these two observations as similar emotional responses. Many of the survivors get through the interview process without emotional breakdown. Of course, those who cry or sob do not start to cry immediately at the beginning of the interview; it takes time until the moment of crying arrives. This leads to the question, overall are women or men more prone to cry? Which group is more likely to start to cry earlier throughout the interview? If they cry, how long does it take for them to start to cry for the first time?

To answer these questions, we modelled the probability of the event not crying in the function of the interview time. We focused on the event not taking place, and included both those who cry or sob at least once and those who do not cry or sob at all (see the censoring process and the survival function in the Methods section). With this focus, we gathered insights into whether men or women tend to collect themselves for a longer period of time (Figure 1). Our results show that women are significantly less likely to collect themselves and more prone to experience crying and sobbing (p-value = 0.0226) earlier than men (Figure 2, see the description of the Hazard ratio and the Hazard function and the associated significance tests in the Methods section). Finally, we also modelled the waiting time for a man or woman to start crying in those interviews where these events actually occur. The result is in line with the previous finding. We found that the waiting time for women to start to cry or sob tends to be significantly shorter (see the description of the use of the Gamma distribution with Alpha=1 in the Methods section).

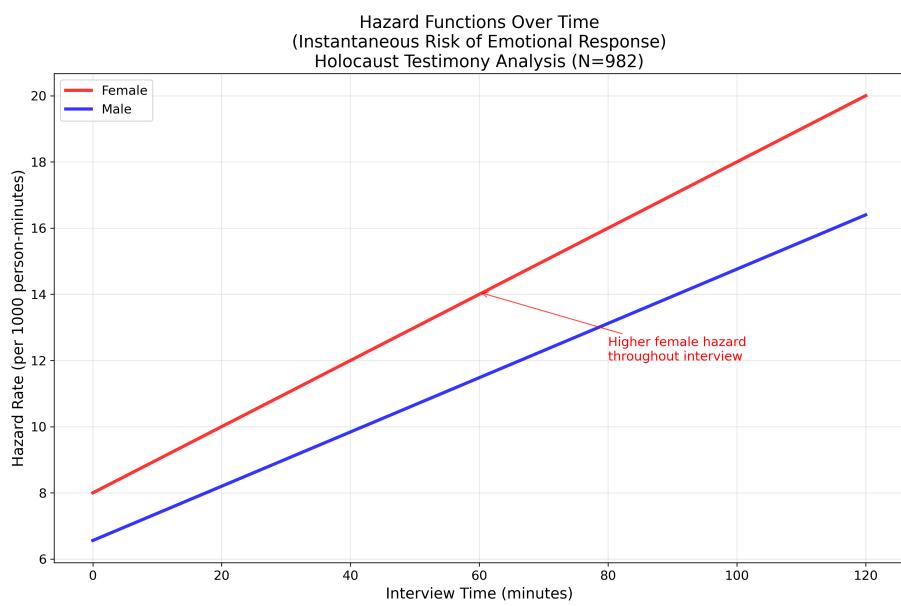
### 4 Results: single layer and multiple crying and sobbing events

In the previous section, we focused on the waiting time until the first event. Nevertheless, those who cry in interviews are likely to cry multiple times. This leads to a number of questions. Is the mean number of crying events more for women or men in the interviews we have? What is the waiting time for these multiple events to occur for the two genders? In addition, we can also ask whether the crying events tend to cluster in close proximity, that is to say, whether they follow each other with short intervals? Alternatively, do they tend to set apart throughout the interview process?

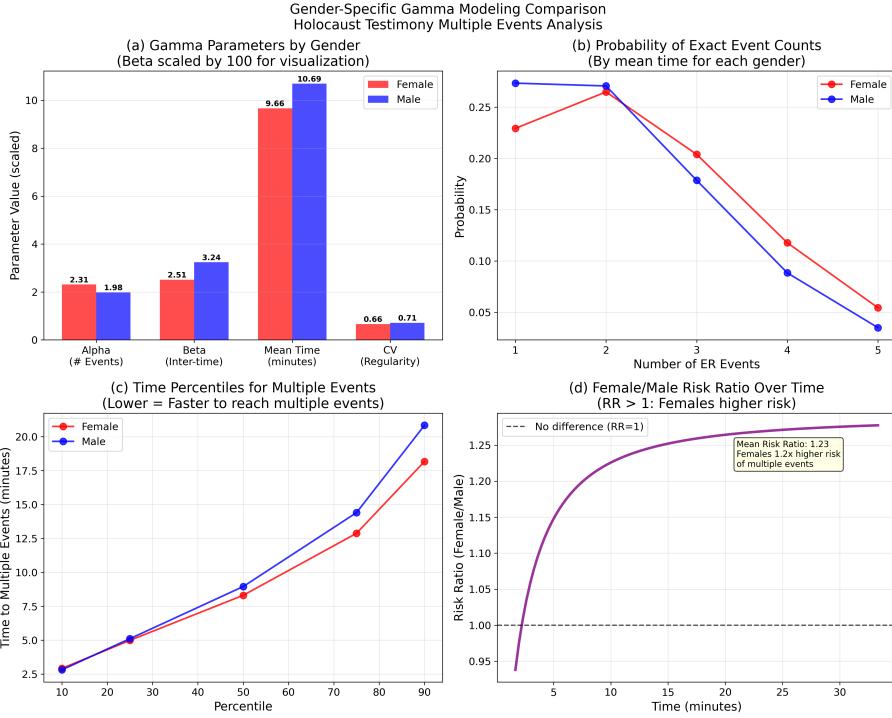
Overall, the probability of multiple crying events (equal or more than 3) is low in our data set; on average survivors cry around two times through the interview process (Figure 3B). However, we found that the mean number of crying events in women testimonies is higher (Figure 3A); according to our analysis, the difference is statistically significant (see Figure 3D, as well as the use of Gamma distribution for modelling multiple events along with the Female / Male risk ratio analysis for multiple events in the Methods section). Our analysis also demonstrates that women reach the mean number of multiple events quicker than men (see Figure 3C). Finally, in our data set, crying events by women tend to cluster in closer proximity than those by men. (see the waiting time between crying events estimated with the Beta parameter of Gamma distribution on Figure 3A).



**Figure 1:** Probability that survivors have not yet cried during the interview (survival function  $S(t)$ )



**Figure 2:** Modeling the instantaneous risk (hazard) of a first crying event by gender



**Figure 3:** Modeling the waiting time to multiple crying events in testimonies of women and men using the Gamma distribution

## 5 Results: two layers and single crying and sobbing event

In the last three decades, Holocaust scholars have studied gender differences in Nazi concentration camps, where men and women were usually detained in different sectors. Scholarship has pointed out systematic differences between experiences by women and men [31]. For instance, it has been argued that solidarity was stronger among women. This and other similar findings of previous scholarship make us hypothesize that women and men cry when recalling different types of memories and concentration camp experiences.

To test this hypothesis, we applied the multilayered approach outlined above. Again, the sequence of memories and experiences discussed through the interview process is one layer. In our data set, memories and experiences are encoded into topics from a purpose-built thesaurus; each interview in the data set is available as a time-stamped and minute-based sequence of segments; each segment is annotated with the topics discussed in the segment (see the data description and the details of the topical annotation in the Appendix). The other layer is given by the time-stamped moments of crying or sobbing. The availability of these two timestamped and therefore parallel layers enable us to inquire whether there are different topics triggering crying or sobbing for women and men - with delay and without delay.

To answer this question, in this study we worked with higher level topics only. As described in the Appendix, the thousands of topics used to annotate interview segments derive from a purpose-built and patented thesaurus or controlled vocabularies. This thesaurus, developed by our data provider, the USC Shoah Foundation, is structured as a directed acyclic graph. For this study, each low level and highly specific topic was replaced with the highest topics in the hierarchy, which are the core nodes of the graph structure. In total, we worked with twenty-three single topics and studied which one triggers a crying event for men and women (see the description of odds ratio analysis in the Methods section).

We found that the following topics are significantly likely (with p-value below 0.05 with

multiple-tests correction) to trigger crying if both men and women are included: captivity, mistreatment and death, discussion of religion and philosophy, forced labour. Nonetheless, with gender included the results are different. There is a statistically significant probability that women start to cry at the discussion of forced labour. By contrast, captivity, mistreatment, discussion of religion and philosophy are associated with men crying. It is important to underline, there are also other topics likely to trigger crying but they lack statistical significance. Generally speaking, our results related to interrelationship between the verbal layer of topics and the non-verbal layer of crying remains preliminary; we need further post-processing of the topical thesaurus to work with lower level topics.

## 6 Method description

### 6.1 Study design and case study mapping

Let  $\mathcal{I} = \{1, \dots, N\}$  index oral-history interviews. For interview  $i \in \mathcal{I}$ , let  $D_i > 0$  denote the observed interview duration (in continuous time units), and let  $\{t_{i,1} < \dots < t_{i,k_i}\} \subset (0, D_i]$  be the time stamps at which an *emotional breakdown* is annotated (crying or sobbing). Define the counting process

$$N_i(t) = \sum_{j=1}^{k_i} \mathbf{1}\{t_{i,j} \leq t\}, \quad t \in [0, D_i],$$

and the at-risk process  $Y_i(t) = \mathbf{1}\{t \leq D_i\}$  [16; 25]. We study:

1. **Time to first breakdown:**  $T_i = \inf\{t > 0 : N_i(t) \geq 1\}$  if  $k_i \geq 1$  and  $T_i = D_i$  otherwise, with event indicator  $\delta_i = \mathbf{1}\{k_i \geq 1\}$ .
2. **Recurrent breakdowns:** the trajectory  $\{N_i(t), 0 \leq t \leq D_i\}$  and gap times  $W_{i,j} = t_{i,j} - t_{i,j-1}$ , with  $t_{i,0} := 0$ .

Interviews without a breakdown are *administratively right-censored* at  $D_i$ . We assume independent interviews and non-informative censoring (conditional on the observed history, interview termination is independent of future breakdowns) [15; 16]. In the case study,  $N = 982$  interviews (women = 454, men = 528) are analyzed. Gender is  $G_i \in \{0, 1\}$  (0=female, 1=male). When thematic annotation is available, the current topic is a time-varying covariate  $Z_i(t) \in \{1, \dots, K\}$  (piecewise-constant over topic segments).

### 6.2 Nonparametric estimands

For time to first breakdown, the survival and cumulative hazard are

$$S(t) = \Pr(T > t), \quad H(t) = \int_0^t h(u) du,$$

with  $h(t)$  the hazard. We estimate  $S(t)$  by Kaplan–Meier and  $H(t)$  by Nelson–Aalen [14; 16; 19]. Group contrasts (men vs. women) use the log-rank test [29]. As an assumption-lean summary, we report the restricted mean survival time (RMST) up to  $\tau_0$ ,

$$\text{RMST}(\tau_0) = \int_0^{\tau_0} S(t) dt, \quad \Delta\text{RMST} = \text{RMST}_{\text{men}}(\tau_0) - \text{RMST}_{\text{women}}(\tau_0),$$

with inference following [22; 30].

### 6.3 Semiparametric hazard regression (first breakdown)

Cox proportional hazards (PH) models quantify covariate effects [6]:

$$h_i(t) = h_0(t) \exp\{\beta_G G_i + \gamma^\top x_i + \eta^\top z_i(t)\},$$

where  $h_0$  is an unspecified baseline,  $x_i$  are time-invariant covariates, and  $z_i(t)$  are time-varying covariates (e.g., topic indicators). Parameters are estimated by the partial likelihood

$$L(\theta) = \prod_{j=1}^J \frac{\exp\{\theta^\top X_{(j)}\}}{\sum_{i \in \mathcal{R}(t_{(j)})} \exp\{\theta^\top X_i(t_{(j)})\}},$$

with distinct event times  $t_{(1)} < \dots < t_{(J)}$ , risk set  $\mathcal{R}(t)$ ,  $X_i(t) = (G_i, x_i, z_i(t))$ , and  $\theta = (\beta_G, \gamma, \eta)$ . Because covariates vary over time, we compute cluster-robust (sandwich) standard errors with interview as the clustering unit [15; 18]. PH diagnostics use scaled Schoenfeld residuals and global tests; if PH is violated, we include time-interactions or rely on RMST contrasts [11; 22; 24; 30].

### 6.4 Recurrent breakdowns: counting-process models

For multiple breakdowns within an interview, the intensity is

$$\lambda_i(t \mid \mathcal{H}_i(t)) = Y_i(t) \lambda_0(t) \exp\{\beta_G G_i + \gamma^\top x_i + \eta^\top z_i(t)\},$$

with history  $\mathcal{H}_i(t)$ . We fit (i) Andersen–Gill (AG) with robust variance [3]; (ii) Prentice–Williams–Peterson (PWP) gap-time, which resets the clock after each event [21]; and (iii) Wei–Lin–Weissfeld (WLW) marginal models with event-order strata and interview-level robust variance [32]. We also plot the cumulative mean function of events over time (nonparametric MCF) [5; 19].

**Parametric interpretation (waiting time to the  $k$ th event).** Under a homogeneous Poisson breakdown process with rate  $\lambda_g$  in group  $g$ , the waiting time to the  $k$ th event is

$$T^{(k)} \sim \text{Gamma}(\alpha = k, \text{rate} = \lambda_g), \quad \mathbb{E}[T^{(k)}] = \frac{k}{\lambda_g}.$$

For  $k = 1$  this reduces to the Exponential distribution ( $\alpha = 1$ ). We use these fits descriptively (by gender), while inference relies on the semiparametric models [5].

### 6.5 Topic-specific hazards and multilayer alignment

We represent each interview in start–stop form with one record per maximal segment of constant covariates. Let  $\{[s_{i,\ell}, e_{i,\ell})\}_{\ell=1}^{L_i}$  be the segmentation; within segment  $\ell$ ,  $z_i(t) \equiv z_{i,\ell}$  and the likelihood contribution occurs at segment endpoints with events. Topic-specific effects are obtained via topic indicators in  $z_i(t)$  or by stratifying on  $Z_i(t)$ ; interactions  $G_i \times Z_i(t)$  allow gendered differences by topic. Multiple testing across topics is addressed with Bonferroni and FDR control [1; 4].

### 6.6 Rate–ratio models for total event counts (robustness)

As a coarse, time-agnostic robustness check for recurrence, we model the *number* of breakdowns per interview,  $K_i = N_i(D_i)$ , using Poisson or negative-binomial regression with a duration offset:

$$\log \mathbb{E}[K_i \mid G_i, x_i] = \alpha + \beta_G G_i + \gamma^\top x_i + \log D_i,$$

reporting the gender *rate ratio*  $\exp(\beta_G)$  and its confidence interval [13]. This complements the event-history analyses.

## 6.7 Topic-trigger association via odds ratios (time-agnostic)

To align with the descriptive “trigger” analyses in the Results, we also compute odds ratios (ORs) for whether a crying/sobbing event occurs when (or shortly after) a high-level topic is present. For a given topic and gender, construct a  $2 \times 2$  table of segments: topic present/absent  $\times$  event observed/not observed within a pre-specified window (e.g., same minute or within  $L$  minutes). The odds ratio and large-sample CI are

$$\text{OR} = \frac{ad}{bc}, \quad \log(\text{OR}) \pm 1.96 \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}},$$

with Fisher’s exact test  $p$ -values when counts are sparse; multiplicity across topics is controlled by FDR [1; 4; 28]. These ORs are descriptive and complement the time-to-event models.

## 6.8 Estimation, uncertainty, and reporting

Point estimates are accompanied by 95% confidence intervals: Greenwood-type for KM, robust sandwich for Cox/AG/WLW/PWP, delta-method for RMST, exact or Wald intervals for ORs, and robust SEs for count models [1; 13; 15; 16]. Unless noted, we use two-sided  $\alpha = 0.05$  with multiplicity adjustments as above.

## 6.9 Relevance to the case study

Within this framework we address: (i) how long witnesses typically speak before a first breakdown and whether this differs by gender/topic (KM/NA, RMST, Cox); (ii) how the instantaneous risk evolves over an interview and across topics (Cox with time-varying covariates; AG/PWP/WLW); (iii) whether total breakdown rates differ by gender (count models); and (iv) which topics are associated with breakdowns in a time-agnostic sense (ORs). The Gamma summary provides an interpretable parametric benchmark for the waiting time to the  $k$ th breakdown [5; 16; 25].

## 7 Discussions and future research

Time is an essential component of oral history interviews. To put it simply, as the interview progresses, it unfolds over time. Analyzing the temporal aspects of interviews in reference to different layers of multimodality is much needed. In this paper, we showed how survival analysis can support these analytical efforts. Most importantly, we demonstrated how survival analysis enables the generalization of temporal observations over a large number of interviews divided according to gender. With this approach, we could compare how female and male survivors of the Holocaust recall their traumatic past in their belated oral history interviews. Our results indicate systematic differences between men and women. By focusing on the temporal aspects of the event of crying, we found solid evidence that men and women tend to cry at different times and after recalling different types of Holocaust experiences. Nevertheless, survival analysis has a number of limitations, which pinpoint the need for further research.

First, survival analysis is a tool to analyze a given data set. It does not enable us to go beyond the data available and study the underlying population. In our case, this means that our results remain valid in reference to the data set we have. However, to go beyond the generalizations and evidence inferred from the data and to talk about the entire population of survivors, we need further statistical tools such as the ones provided by the Bayesian framework and unseen species studies [9]. Currently, our results are only indicative in relation to the entire survivor population.

Second, survival analysis works with time between two distinct events such as for instance the beginning of an interview and the moment one starts crying. Nevertheless, between these two

events, there can be a number of other events influencing the temporal dynamics of the interview process. Hence, we need parallel sequence analysis to complement survival analysis.

Third, survival analysis helps explore events that follow each other with delay. In this paper, we treated this “following each other” as triggering. We applied statistical significance testing to check whether “following each other” is just a coincidence (null hypothesis) or whether we can assume causality. To which extent the rejection of null hypothesis is indeed an indicator of true causality is a question of philosophical discussions [8]. This paper is not to address this problem; however, one has to see our results with keeping these discussions in mind.

To summarize, survival analysis is a valuable tool to support multimodal data analysis but further statistical tools are needed to achieve more comprehensive results.

## Code availability

The analysis scripts and reproducible code used in this study are available on GitHub at <https://github.com/toth12/multimodal-survival-analysis>. The repository includes data-processing pipelines, modeling code (survival and recurrent-event analyses), figure-generation scripts, and instructions to reproduce all results from a clean environment.

## Acknowledgements

This paper was generously funded by the Luxembourg National Research Fund and the German Research Foundation (grant number: C24/ID/18896236/VOICES). We are thankful to the USC Shoah Foundation for the opportunity to work with their testimonies. Finally, we are grateful to the anonymous reviewers for their thorough comments and suggestions.

## References

- [1] Agresti, Alan. *Categorical Data Analysis*. 3rd ed. Hoboken, NJ: Wiley, 2013.
- [2] Amorim, Leila DAF and Cai, Jianwen. “Modelling recurrent events: A tutorial for analysis in epidemiology”. In: *International Journal of Epidemiology* 44, no. 1 (2014), pp. 324–333.
- [3] Andersen, Per K. and Gill, Richard D. “Cox’s regression model for counting processes: a large sample study”. In: *The Annals of Statistics* 10, no. 4 (1982), pp. 1100–1120. DOI: [10.1214/aos/1176345976](https://doi.org/10.1214/aos/1176345976).
- [4] Benjamini, Yoav and Hochberg, Yosef. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 57, no. 1 (1995), pp. 289–300.
- [5] Cook, Richard J. and Lawless, Jerald F. *The Statistical Analysis of Recurrent Events*. New York: Springer, 2007.
- [6] Cox, D. R. “Regression Models and Life-Tables”. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 34, no. 2 (1972), pp. 187–220.
- [7] Dash, Amulya and Sharma, Yashvardhan. “A collaborative approach to multimodal machine translation: VLM and LLM”. In: *Proceedings of the 17th International Conference on Agents and Artificial Intelligence*. SCITEPRESS - Science and Technology Publications, 2025, pp. 1412–1418. DOI: [10.5220/0013379900003890](https://doi.org/10.5220/0013379900003890).
- [8] Efron, Bradley. “The Efficiency of Cox’s Likelihood Function for Censored Data”. In: *Journal of the American Statistical Association* 72, no. 359 (1977), pp. 557–565. DOI: [10.1080/02640414.2012.753636](https://doi.org/10.1080/02640414.2012.753636).

- [9] Efron, Bradley and Thisted, Ronald. “Estimating the number of unseen species: How many words did Shakespeare know?” In: *Biometrika* 63, no. 3 (1976), p. 435. DOI: 10.2307/2335721.
- [10] Felman, Shoshana and Laub, Dori. *Testimony: Crises of witnessing in literature, psychoanalysis, and history*. Taylor & Francis, 1992.
- [11] Grambsch, Patricia M. and Therneau, Terry M. “Proportional Hazards Tests and Diagnostics Based on Weighted Residuals”. In: *Biometrika* 81, no. 3 (1994), pp. 515–526.
- [12] Greenspan, Henry. “From testimony to recounting”. In: *Beyond Testimony and Trauma*. University of British Columbia Press, 2015, pp. 141–169. DOI: 10.59962/9780774828949-007.
- [13] Hilbe, Joseph M. *Negative Binomial Regression*. 2nd ed. Cambridge: Cambridge University Press, 2011.
- [14] Kaplan, E. L. and Meier, Paul. “Nonparametric estimation from incomplete observations”. In: *Journal of the American Statistical Association* 53, no. 282 (1958), pp. 457–481. DOI: 10.1080/01621459.1958.10501452.
- [15] Klein, John P., Houwelingen, van, Ibrahim, Joseph G., and Scheike, Thomas H. *Handbook of survival analysis*. CRC Press, 2016.
- [16] Klein, John P. and Moeschberger, Melvin L. *Survival Analysis: Techniques for Censored and Truncated Data*. 2nd ed. Springer, 2003. DOI: 10.1007/b97377.
- [17] Langer, Lawrence L. *Holocaust testimonies: The ruins of memory*. Yale University Press, 1991.
- [18] Lin, D. Y. and Wei, L. J. “The Robust Inference for the Cox Proportional Hazards Model”. In: *Journal of the American Statistical Association* 84, no. 408 (1989), pp. 1074–1078.
- [19] Nelson, Wayne. “Confidence Limits for Recurrence Data—Applied to Cost or Other Cumulative Functions”. In: *Technometrics* 37, no. 2 (1995), pp. 147–158.
- [20] Pollin-Galay, Hannah. *Ecologies of Witnessing: Language, place, and holocaust testimony*. Yale University Press, 2018.
- [21] Prentice, Ross L., Williams, Barbara J., and Peterson, Allan V. “On the Regression Analysis of Multivariate Failure Time Data”. In: *Biometrika* 68, no. 2 (1981), pp. 373–379.
- [22] Royston, Patrick and Parmar, Mahesh K. B. “The Use of Restricted Mean Survival Time to Estimate the Treatment Effect in Randomized Clinical Trials When the Proportional Hazards Assumption Is in Doubt”. In: *Statistics in Medicine* 30, no. 19 (2011), pp. 2409–2421.
- [23] Saltzman, Lisa and Rosenberg, Eric M. *Trauma and visuality in modernity*. UPNE, 2006.
- [24] Schoenfeld, David. “Partial Residuals for the Proportional Hazards Regression Model”. In: *Biometrika* 69, no. 1 (1982), pp. 239–241.
- [25] Singer, Judith D. and Willett, John B. *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*. New York: Oxford University Press, 2003.
- [26] Smits, Thomas and Wevers, Melvin. “A multimodal turn in Digital Humanities. Using contrastive machine learning models to explore, enrich, and analyze digital visual historical collections”. In: *Digital Scholarship in the Humanities* 38, no. 3 (2023), pp. 1267–1280. DOI: 10.1093/11c/fqad008.
- [27] Smyth, Hannah K, Nyhan, Julianne, and Flinn, Andrew. “Exploring the possibilities of Thomson’s fourth paradigm transformation—The case for a multimodal approach to digital oral history?” In: *Digital Scholarship in the Humanities* 38, no. 2 (2023), pp. 720–736. DOI: 10.1093/11c/fqac094.

- [28] Szumilas, Magdalena. “Explaining odds ratios - PMC”. In: *Journal of the Canadian Academy of Child and Adolescent Psychiatry* 19, no. 3 (2010), pp. 227–229.
- [29] Tang, Xinyu. “Log-Rank test”. In: *Handbook for Clinical Research*. New York, NY: Springer Publishing Company, 2014. DOI: 10.1891/9781617050992.0041.
- [30] Uno, Hajime, Claggett, Michael, Cai, Tianxi, Tian, Lu, Schrag, Dave, Buyse, Marc, and Wei, L. J. “Moving Beyond the Hazard Ratio in Quantifying the Between-Group Difference in Survival Analysis”. In: *Journal of Clinical Oncology* 32, no. 21 (2014), pp. 2380–2385.
- [31] Waxman, Zoë. *Women in the Holocaust: A feminist history*. Oxford University Press, 2017. ISBN: 9780191090707.
- [32] Wei, Lee-Jen, Lin, D. Y., and Weissfeld, Lawrence. “Regression Analysis of Multivariate Incomplete Failure Time Data by Modeling Marginal Distributions”. In: *Journal of the American Statistical Association* 84, no. 408 (1989), pp. 1065–1073.

## A Data descriptions

Our data is a collection of 982 oral history testimonies (interviews) with 454 women and 528 men. The interviews were conducted by the USC Shoah Foundation between 1994 and 1998. Trained human annotators of the Foundation annotated each interview with topical keywords from a controlled vocabulary. The thousands of explicitly defined keywords in the controlled vocabulary aim to cover all possible Holocaust experiences and they were developed by the Foundation itself. The annotation process and the controlled vocabulary are patented by Samuel Gustman, Chief Technology Officer of the Foundation: “Method and Apparatus for Cataloguing Multimedia Data,” U.S. Patent 5,832,495 (Nov. 3, 1998).

Specifically, in our data set each interview is available in three different formats: timestamped transcripts with multimodal clues (crying, sobbing, etc) in text format; timestamped topical annotation grouped into minute-based segments in XML; original audio/video files in MP3. The multimodal clues were added throughout the transcription process by the Foundation.

The dataset can be consulted on the website of the Foundation:

<https://vha.usc.edu/>