Playtime Measurement with Survival Analysis

Markus Viljanen, Antti Airola, Jukka Heikkonen, Tapio Pahikkala

***Abstract*—Maximizing product use is a central goal of many businesses, which makes retention and monetization two central analytics metrics in games. Player retention may refer to various duration variables quantifying product use: total playtime or session playtime are popular research targets, and active playtime is well-suited for subscription games. Such research often has the goal of increasing player retention or conversely decreasing player churn. Survival analysis is a framework of powerful tools well suited for retention type data. This paper contributes new methods to game analytics on how playtime can be measured using survival analysis. These methods include visualizations, metrics and an AB-test based on the survival curve. All these methods work on censored data and enable computation of confidence intervals. This is especially important in time and sample limited data which occurs during game development. Throughout this paper, we illustrate the application of these methods to the development of the Hipster Sheep mobile game.**

***Index Terms*—game analytics, survival analysis, playtime, retention.**

I. Playtime in Games

GAME analytics is becoming increasingly important in understanding player behavior [1]. Widespread adoption of games, Internet connectivity and new business models have resulted in data gathering in an unprecedented scale. With increasing availability of data, academia and industry alike are motivated to gain insight into the data through game analytics.

Focal point of analytics is player retention and churn [2]. Retention has been used in connection with many related measures and methods aiming to increase the length of product use [3]–[12]. Better retention simply means players are engaged with the game for longer. Player churn [13]–[20], meaning players quitting the game either momentarily or definitely, decreases product use and is therefore a counterpart of retention. Retention metrics are popular because they are thought to reflect player enjoyment, and increased product use provides increased possibilities for monetization in free-to-play and subscription based games. Game success may be attributed to the process of acquiring new users and retaining these users with effective monetization [2].

In this paper we study what kind of methods can be used to analyze player retention and churn in a timely and effective manner. We describe survival analysis methods that allow analyzing player retention as a duration variable. Survival analysis is well-suited for retention analysis, because it has been developed specifically for duration data which may be censored and highly non-normal. We limit our investigation to understanding, measuring and comparing total playtime. This restriction to playtime allows increased brevity and clarity, however these tools directly extend to any duration data, such as sessions, level progression or subscription time. In addition, survival analysis also accommodates several regression formulations, of which the Cox proportional hazards regression has been the most popular both in games [21], [22] and in standard survival analysis [23]. The research data comes from Tribeflame Ltd.' free-to-play mobile game named Hipster Sheep. The findings of our research enable game developers, managers and publishers to better benchmark their games.

M. Viljanen, A. Airola, J. Heikkonen, and T. Pahikkala are with the Department of Future Technologies, University of Turku, 20014 Turku, emails: majuvi@utu.fi, ajairo@utu.fi, jukhei@utu.fi, aatapa@utu.fi

The rest of the paper is structured as follows. Section II reviews literature introducing retention as a duration and the survival analysis of durations. Section III introduces research objectives and research data. Sections IV-VII provide the research results and analysis. Finally, Section VIII concludes our findings.

II. Literature Review

A. Retention as a Duration

User behavior in terms of measuring duration data has been researched in game analytics [24], [25] and game networking [26]–[33] as a topic of itself. Game analytics literature attempts to understand user retention and how the game itself contributes to it. Game networking often analyzes both active and inactive durations. As a field it is principally concerned with how network quality and related factors add to user retention and how user activity on the other hand imposes a load on servers which operator might try to mitigate. Total playtime [21], [24], [25], session playtime [26]–[33] and session inter-arrival time [26], [27], [29] or idle time, have been popular measurements. Session is commonly defined as a duration of continuous play [26]–[33] but has also been used to refer to a completed match [3]–[8]. Popular retention measurements in long-term games are subscription times [13]–[15] or active periods over calendar time [16]–[20], possibly combined [14]. Session counts [3], [4] and progression [9] are also instances of user retention, which may be modeled as a discrete duration variable.

Based on the literature, there are several candidate variables for measuring retention as a duration:

1) Total playtime
2) Session playtime
3) Total progression
4) Total active or subscription time (MMORPGs etc.)

Total playtime is the total time spent playing the game, in seconds for example. Session playtime corresponds to the duration of continuous play. Total progression relies on a game developer's intuition of how game consumption is transformed into a positive non-decreasing value, a natural example would be levels completed. In games with open-ended goals and long-term gameplay, one may analyze total time active as the calendar time player was engaged in the game world, or total subscription periods such as months until cancellation.

Playtime has been an important focal point in many studies, because it provides a straightforward aggregate metric of total player engagement. A notable example of fitting a parametric distribution is the study of total playtimes [25] in over 3000 Steam games totaling 6 million players which utilized the Weibull distribution for archetype analysis. Other research with parametric models has often investigated exponential [26], [27], [34], Weibull [11], [12], [24], [25], [28], [29], [33], Gamma [24], log-logistic [12], log-normal [21], [24], [33] and Pareto-type [12], [26], [27], [32] distributions.

B. Survival Analysis of Durations

Survival analysis is in the early stages of being applied to game analytics. Some studies have used tools that are central in survival analysis, such as the survival curve of playtimes or the churn rate [28]–[31], where the use of Kaplan-Meier estimate [30], [31] is notable to deal with censored session durations. Studies focusing on survival analysis have researched measuring difficulty with automated playtesting [34], modeling the user process [11], and using the Cox regression extended by time-varying covariates and coefficients [21] or survival ensembles [22]. The first study [21] analyzes total playtime and the second study [22] days active to churn.

The target of survival analysis is a positive duration variable, which is the time to an event of interest. This may be lifetime in demography [35], the time to machine failure in reliability engineering [36] or the time to disease recurrence in medicine [23]. The duration variable may also be discrete, such as lifespan in years or repetitions to failure. The primary reason for the de facto status of survival analysis in many fields is that survival analysis excels with non-normal and censored data, and does not necessitate parametric approaches [23]. Retention exhibits non-normal characteristics: duration is positive, heavily skewed towards zero and often has a long tail. Since in the industry it is often unfeasible to wait until all users have churned to obtain their total playtime, censoring is also present. Furthermore, user retention may not always follow popular parametric models [12], [24], making model-free approaches attractive. The widespread recognition of survival analysis in fields with similar data and the demand for scientific analytics suggests that game analytics will benefit from this approach.

C. Churn and Censoring

Censoring is very common in survival analysis type data. For example, in medical studies patients may drop out of the study before experiencing the event of interest or the study may have a limited follow-up time which terminates the study before every patient has had the event of interest [23]. Such subject is called *censored* and the data is in this case subject to *right censoring*. These subjects contribute information, since the time to failure must be greater than the time to censoring.

In successful games with very long playtimes, censoring is often unavoidable because game developers wish to perform analytics before every user churns. A second important challenge exists in games which is not found in survival analysis: we can always observe if a person contracted a disease or a machine failed, but it is not possible in principle to know if a player has churned definitely. The player may always return to the game; it is only with the passing of time that we increase the confidence this will not happen. This problem manifests in quits without notification because it is sometimes difficult to say whether the user comes back. The problem does not exist if the event is observable. Examples of such events are session end, level failure and subscription cancellation.

The challenge of detecting churn related to total playtime has been dealt with in the literature using various rules to impute churned and non-churned players: assuming players have churned [24], [25] or defining a window of inactivity which implies churn [16]–[19]. Once trained, churn prediction algorithms and user process models could also be used to predict the censoring label [11], [16]–[19]. Nevertheless, none of the current solutions seem perfectly satisfactory as they may add bias depending on the method. Player churn is an extended topic and we further assume that a simple method to impute churn is available. This enables us to focus on the standard methods which are universally applicable [36].

III. Research Objectives and Data

A. Research Focus

In this paper we describe how survival analysis can contribute to playtime analysis. We introduce the following fundamental analyses a game analyst can carry out on data consisting of observed playtimes, using standard survival analysis software such as R [37], SAS and Stata [23], among others.

1) Survival and hazard curves: these two foundational concepts of survival analysis allow studying both visually and analytically the rates at which players churn from the game at different time points. They enable the analyst to better understand the overall quality, including the strong and the weak points of a game.
2) Mean and median provide singular metrics for characterizing the expected and the typical playtime. They allow the analyst to aggregate the data to a single informative number together with confidence intervals.
3) The log-rank test provides a scientific AB-test by comparing the survival curves of different groups (e.g. the players of different versions of a game). This allows the analyst to deduce whether the groups are different to a given degree of confidence.

We chose to use the total playtimes of an in-development mobile game to illustrate every method with a real world survival analysis application. All the presented methods are accompanied with examples from Tribeflame Ltd.'s mobile game named Hipster Sheep. We have further utilized a small sample of 10 of these users to pedagogically demonstrate how each estimate is computed. Following [16], [17], we decided that players playing 14 days within collection time are not churned, leading to censored playtimes.

Fig. 1. Hipster Sheep promotional material also displayed in Google Play, used with the permission of Tribeflame Ltd.

B. Research Data - Playtime Data Set

Hipster Sheep is a commercial grade puzzle game being developed by Tribeflame Ltd. Figure 1 displays promotional material that depicts the game. The game is targeted at young adult females and has an artistic theme of making light fun of hipsters through self-irony. The goal is to guide an anthropomorphic sheep through labyrinths on her or his quest for the next big thing. The game is free-to-play, level-based and uses in-game purchase monetization. Energy mechanics are used to limit the possibility for unlimited free play. Levels combine skill with a hefty dose of luck, as is common in modern free-to-play games.

During the development, there were three significant user acquisition campaigns for versions 1.11, 1.15 and 1.18. Users were purchased through random sampling by advertising in social networks. The purpose was to test the game's appeal in-between successive development cycles. Figure 2 displays daily new users (DNU) and the resulting daily active users (DAU) for these versions. In version 1.11 there were a total of 970 users acquired in early June 2015, in version 1.15 a total of 1246 users were acquired early September 2015 and in version 1.18 a total of 1537 new players arrive, mostly mid-October. The three versions hold 3753 players in total. This excludes those players with only one extremely brief session, since this was deduced to be part of 'acquisition phenomena' rather than gameplay; the game takes a dozen seconds to load.

The total of n playtimes T_i and censoring indicators δ_i form a data set $((T_1, \delta_1), \ldots, (T_n, \delta_n))$. If $\delta_i = 1$, the player has not churned and the playtime is greater or equal to T_i, and $\delta_i = 0$ implies that the churn event occurred at T_i. In Table I, we have randomly sampled a subset of 10 players from version 1.18

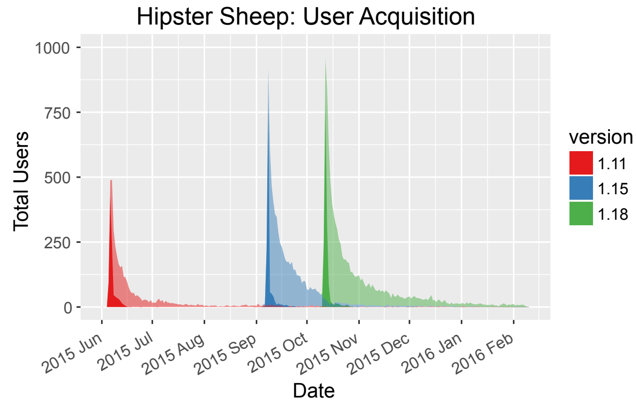

Fig. 2. User acquisition in versions 1.11, 1.15 and 1.18. The DNU is highlighted in dark color and the DAU in transparent color. Each acquisition spans few days, with the resulting user activity diminishing over time.

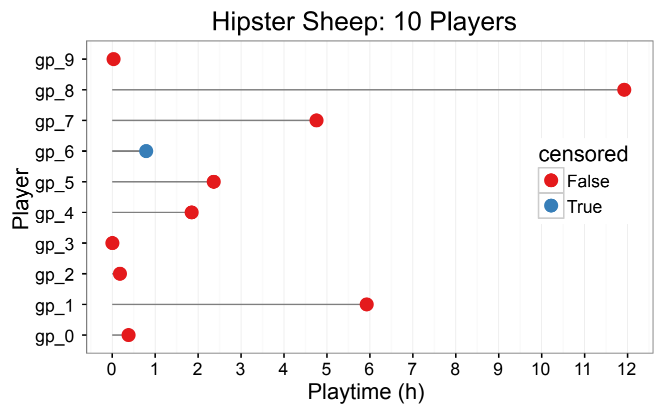

Fig. 3. A random sample of 10 players from Hipster Sheep version 1.18 for Android. Player with identifier 'gp_6' had been playing very recently and was determined to be active, whereas others had churned with a high likelihood.

for Android, where every player has a playtime duration and a censoring indicator. Figure 3 visualizes this data. We see that there is significant early churn, with 40% churning before 1 hour of gameplay, and a heavy tail with one 12 hour gameplay observation. Other players seem to have more typical 1-6 hour playtimes. One player happened to be censored.

Survival analysis is commonly formulated using a sequence of strictly increasing *distinct* playtimes $t_1 < \ldots < t_m$. At every distinct playtime t_i we count the number of observed playtimes $d_i = \sum_{j=1}^{n} \mathbb{I}(T_j = t_i \wedge \delta_j = 0)$ and the number of censored playtimes $c_i = \sum_{j=1}^{n} \mathbb{I}(T_j = t_i \wedge \delta_j = 1)$. The number of surviving non-censored 'at risk' players is denoted by $n_i = \sum_{j=1}^{n} \mathbb{I}(T_j \geq t_i)$.

IV. Survival Analysis

Survival analysis helps to understand playtime data. In this section, two foundational concepts in survival analysis are introduced: the survival curve and the hazard. These concepts enable the analyst to analyze the playtime data both visually and computationally. The survival curve is a natural way to visualize the proportion surviving of a given population and the hazard function is often motivated as the cause for a given survival curve, which enables simpler analysis.

TABLE I
HIPSTER SHEEP: 10 PLAYERS

Player	Playtime	Censored	Playtime (hours)
gp0	00:22:51	False	0.38
gp1	05:55:32	False	5.93
gp2	00:10:48	False	0.18
gp3	00:00:13	False	0.00
gp4	01:50:59	False	1.85
gp5	02:21:48	False	2.36
gp6	00:47:27	True	0.79
gp7	04:45:25	False	4.76
gp8	11:55:22	False	11.92
gp9	00:01:53	False	0.03

A. The Survival Function

Statistically speaking, the playtimes T_i are a sample of a random variable $T > 0$ of the population playtimes and based on the sample we seek to analyze the distribution of T. In the discrete case, where T is for example playtime rounded to hours, we use a probability mass function (PMF), or the probability of failure at t: $f(t) = \mathbb{P}[T = t]$. If T is continuous, we define a probability density function (PDF) instead:

$$f(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}[t \leq T < t + \Delta t]}{\Delta t}. \quad (1)$$

Regardless whether we used the PMF or the PDF, a cumulative density function (CDF) is used to describe the accumulated probability of playtime less or equal to t. In survival analysis, we often analyze a survival function (SF) [36] instead, which gives the probability of having playtime greater than t:

$$\begin{aligned} F(t) &= \int_0^t f(u)du = \mathbb{P}[T \leq t] \\ S(t) &= 1 - F(t) = \mathbb{P}[T > t]. \end{aligned} \quad (2)$$

As a complement to the cumulative density function, the survival function is a monotonously decreasing function and has the property that $S(0) = 1$ and $S(t) \to 0$ as $t \to \infty$.

B. The Hazard Function

Survival analysis distributions are often easiest to understand in terms of geometric decay of players. Since churned players are no longer at risk of churn, it is useful to contemplate constant churn rate acting on the remaining players. We might also have churn rates with high initial churn, and thereon simple constant churn. For example, in Table II, 50% of players play more than one session, but after the second session 80% survive to play session 3, of those 80% survive to play session 4, etc. Churn in this case refers to the number of players who after session i do not play the next session $i + 1$ and survival quantifies the number of players playing more than the ith session.

This is formalized in the concept of a hazard function [36]. For the discrete case, the hazard function quantifies the proportion of the remaining players who churn:

$$h(t) = \mathbb{P}[T = t \mid T \geq t] = \frac{f(t)}{S(t-1)}. \quad (3)$$

TABLE II
1000 PLAYERS CHURN EXAMPLE

Function	Session 1	Session 2	Session 3	Session 4	…
Retention rate $r(t)$	50%	80%	80%	80%	…
Churn rate $h(t)$	50%	20%	20%	20%	…
# Failing $f(t)$	500	100	80	64	…
# Surviving $S(t)$	1000	500	400	320	…

For the continuous case, the hazard is the instantaneous failure rate in the remaining player base, motivated as the limit:

$$h(t) = \lim_{\Delta t \to 0} \frac{\mathbb{P}[t \leq T < t + \Delta t \mid T \geq t]}{\Delta t} = \frac{f(t)}{S(t)}. \quad (4)$$

An important point to note is that the continuous hazard is not the probability of failure at time t, it can be greater than 1. An approximation in a small interval for the probability of failure is $\mathbb{P}[t \leq T < t + \Delta t \mid T \geq t] \approx h(t)\Delta t$. The relationship of proportional failure probability and rate is analogous to that of the PMF and the PDF. These two settings can actually be treated together with the Riemann-Stieltjes integral, for more information we refer the reader to [36], [38].

C. Connection of Hazard and Survival

Given the hazard or the survival function, one can derive the other. In practical applications the hazard is often analyzed for simpler interpretations and the survival curve derived as a function of the hazard. For the discrete case it is easy to see that a product formulation is possible: at point t, the survival is the product of the fractions remaining after churn events at $u = 1, 2 \ldots, t$. For example, in Table II the survival after session three is: $S(3) = (1 - 50\%)(1 - 20\%)(1 - 20\%) = 32\%$, i.e. 320 players.

$$S(t) = \prod_{u=1}^{t} [1 - h(u)]. \quad (5)$$

In the continuous case, we take a product integral which works like the Riemann integral in partitioning the domain. Instead of a sum, the result is the limit of a product of terms over the partition, consisting of surviving fractions $1 - h(t)\Delta t$. In fact, the less well-known product integral may be written in terms of the Riemann integral by taking the logarithm [38]:

$$S(t) = \prod_{(0,t]} [1 - h(u)\,du] = \exp\left[-\int_0^t h(u)\,du\right]. \quad (6)$$

The integral of the hazard is the cumulative hazard function [36]. Its utility is explained in how a proportional change in $S(t)$ corresponds to a linear change in $H(t)$. Taking the logarithm gives the cumulative hazard in terms of the survival function:

$$H(t) = \int_0^t h(u)\,du = -\log[S(t)]. \quad (7)$$

In the simplest case, the hazard $h(t) = \lambda$ is homogeneous implying that the churn rate is constant over time. This hazard can be used to derive two well-known distributions: geometric distribution for the discrete case and exponential distribution for the continuous case. Of the common survival distributions listed in Table IV, Weibull is one of the most popular [39], and it also has wide applicability in games [11], [12], [25].

V. Playtime Survival

In this section, the survival and the hazard function are utilized to measure game goodness. First the theory is introduced using the 10 player sample and then the methods are applied to the game data set. In distribution fitting one has the problem of choosing a parametric model. However, in survival analysis it is not actually necessary to guess distributions; the data over the follow-up time can be used to make model free estimates. These are called nonparametric methods [38].

A. Fitting a Parametric Survival Model

Suppose that one has a reason to believe that data follows a parametric model and the hazard or the survival is specified. What next? Fitting a distribution is often done utilizing the maximum likelihood (ML) theory [39]. Specifically, for duration observations $T_1, \ldots, T_n$ and censoring indicators $\delta_1, \ldots, \delta_n$, a PDF/PMF $f(t \mid \theta)$ parametrized by θ is fitted by assuming the observations i.i.d. and finding the parameters θ^* which maximize the likelihood $L(\mathcal{D}, \theta)$ of observed data:

$$L(\mathcal{D}, \theta) = \prod_{i=1}^{n} f(T_i \mid \theta)^{1-\delta_i} S(T_i \mid \theta)^{\delta_i} \tag{8}$$

$$\theta^* = \operatorname*{argmax}_{\theta} l(\mathcal{D}, \theta). \tag{9}$$

The logarithm of the likelihood $l(\mathcal{D}, \theta) = \log[L(\mathcal{D}, \theta)]$ is taken in practice to avoid numerical errors associated with extremely small quantities. The ML estimate may be found iteratively using optimization algorithms such as Newton-Raphson [38].

For example, to fit the exponential distribution in Figure 4. one minimizes the likelihood in terms of the objective

$$L(\mathcal{D}, \theta) = \prod_{i=1}^{n} (\lambda e^{-\lambda T_i})^{1-\delta_i} (e^{-\lambda T_i})^{\delta_i} = \lambda^d e^{-\lambda R}, \tag{10}$$

where we have defined the number of observed churns $d = \sum_{i=1}^{n}(1-\delta_i)$ and the total time at risk of churning $R = \sum_{i=1}^{n} T_i$. The log-likelihood is maximized when the derivative is zero. In this case we can directly find the ML estimate:

$$l'(D, \lambda^*) = \frac{d}{\lambda^*} - R = 0 \Rightarrow \lambda^* = \frac{d}{R}. \tag{11}$$

Given the survival times in Table I, there are 10 players with $d = 9$ churning. The total time at risk is the sum of accumulated playtimes: $R = 0.38 + 5.93 + 0.18 + 0.00 + 1.85 + 2.36 + 0.79_+ + 4.76 + 11.92 + 0.03 = 28.21$ (hours). Therefore, we obtain a failure rate or hazard, of $\lambda = \frac{9 \text{ churns}}{28.21 \text{ h}} = 0.32$ churns/h.

95% confidence intervals (C.I.) for the parameter λ^* may be obtained using the normal approximation $X \pm 1.96\sqrt{\text{Var}[X]}$ where 1.96 is the value of standard normal distribution such that $\mathbb{P}[-z \leq Z \leq z] = 0.95$. To estimate confidence intervals for an asymptotically normally distributed quantity one therefore needs an estimate for its variance.

The variance estimate can be obtained by substituting the maximum likelihood parameter into the inverse of the negative Hessian [37], which in the single parameter case equals the second derivative $l''(\mathcal{D}, \lambda) = -d/\lambda^2$:

$$\text{Var}[\lambda] = \left(-(-d/\lambda^{*2})\right)^{-1} = d/R^2. \tag{12}$$

The failure rate estimate with confidence intervals therefore is $\lambda = 0.32 \pm 0.21$ churns/h.

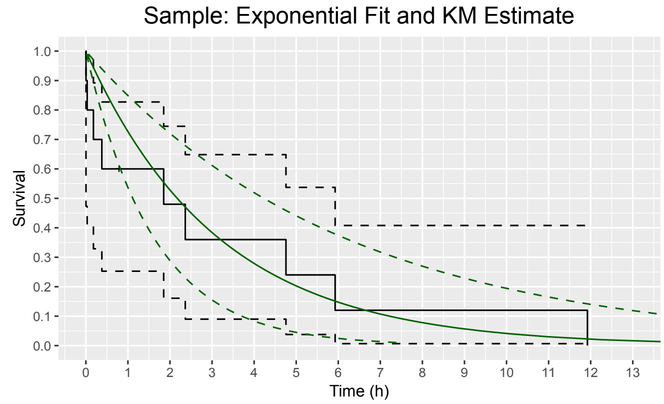

Fig. 4. Exponential fit (green) $\lambda = 0.32$ to the sample of 10 players, with confidence intervals (green, dashed). Contrasted to the KM baseline (black), the early observations may not fit the simple model.

TABLE III
Sample: KM Computation

Time (h)	At risk n_i	Churn d_i	Haz. $\frac{d_i}{n_i}$	Cum. Haz. (NA)	Surv. (KM)	CI.l. 95%	CI.u. 95%
0.00	10	1	0.10	0.10	0.90	0.47	0.99
0.03	9	1	0.11	0.21	0.80	0.41	0.95
0.18	8	1	0.13	0.34	0.70	0.33	0.89
0.38	7	1	0.14	0.48	0.60	0.25	0.83
1.85	5	1	0.20	0.68	0.48	0.16	0.75
2.36	4	1	0.25	0.93	0.36	0.09	0.65
4.76	3	1	0.33	1.26	0.24	0.04	0.54
5.93	2	1	0.50	1.76	0.12	0.01	0.41
11.92	1	1	1.00	2.76	0.00		

B. Fitting a Nonparametric Survival Model

Since a chosen parametric model may not always fit the data, it is often desirable to use an empirical estimate as a baseline. If there are no censored observations, it is straightforward to compute the SF empirically as the fraction of playtimes greater than t: $\widehat{S}(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(T_i > t)$. However, if there are censored observations we need to use a Kaplan-Meier (KM) estimate [23] for an unbiased approximation of the SF:

$$\widehat{S}_{KM}(t) = \prod_{t_i \leq t} \left(1 - \frac{d_i}{n_i}\right). \tag{13}$$

The estimator is easiest to describe with an example. Table III. and Figure 4 show the estimate calculated for the 10 player sample in Table I. At every event time t_i, we compute the remaining fraction $1 - d_i/n_i$ and multiply it with the KM-estimate of survival at the previous failure time $\widehat{S}_{KM}(t_{i-1})$ to obtain the surviving population $\widehat{S}_{KM}(t_i)$. Note how the one censored event time at $t = 0.79$ is not in the table of event times but reduces the risk set at $t = 1.85$.

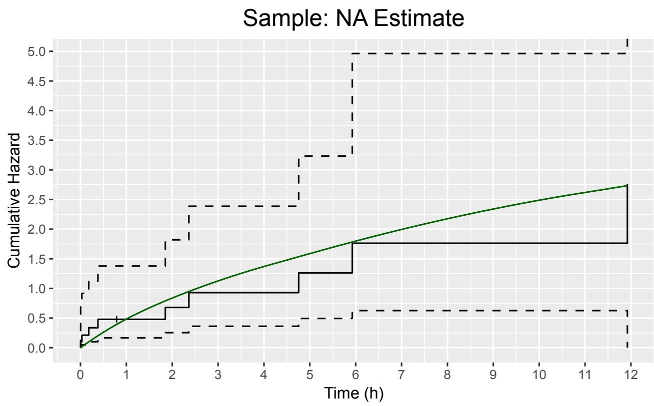

Fig. 5. Nelson-Aalen estimate (black) and confidence intervals (black, dashed) with Epanechnikov kernel smoothing (green) of the cumulative hazard in the small sample. The survival corresponding to this estimate never reaches zero.

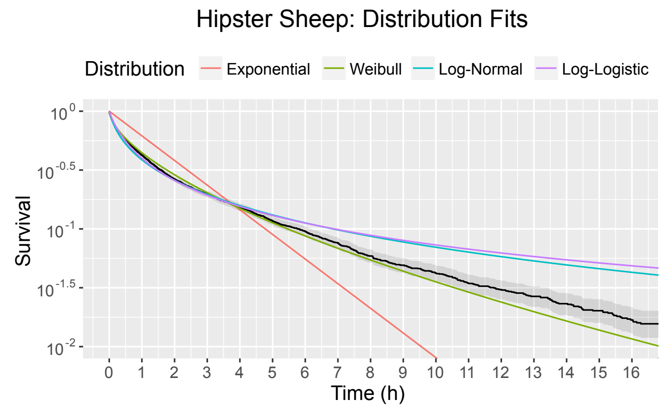

Fig. 6. Kaplan-Meier estimator and distribution fits for the entire population in Hipster Sheep. Confidence in the KM estimate is high due to sample size. The Weibull distribution fits data best with others having significant flaws.

C. Estimating Playtime Cumulative Hazard

An alternative is the Nelson-Aalen (NA) [23] estimate of the cumulative hazard given by the sum of fractions churning:

$$\widehat{H}_{NA}\left(t\right)=\sum_{t_i\leq t}\frac{d_i}{n_i}. \tag{14}$$

A nonparametric hazard estimate [38] requires smoothing the cumulative hazard step function estimate with kernels, for example. Various kernels exists; popular choices are the uniform kernel, the Epanechnikov kernel and the Gaussian kernel. Kernel $K\left(t\right)$ is a mass of density concentrated around zero with a total area of one, spread determined by its bandwidth b. Kernel estimation gives a smooth hazard estimate:

$$\widehat{h}\left(t\right)=\frac{1}{b}\sum_{i=1}^{m}K\left(\frac{t-t_i}{b}\right)\frac{d_i}{n_i}. \tag{15}$$

Of course, using the KM-estimate we can derive a cumulative hazard estimate by $\widehat{H}_{KM}(t)=-\log[\widehat{S}_{KM}(t)]$. Equivalently for the NA-estimate we have $\widehat{S}_{NA}(t)=\exp\left[-\widehat{H}_{NA}(t)\right]$. Both estimators are utilized extensively in practice [38].

It is possible to compute confidence intervals for the KM. The variance may be approximated with the delta method [38]:

$$\mathrm{Var}[\widehat{S}_{KM}(t)]=[\widehat{S}_{KM}(t)]^2\prod_{t_i\leq t}\left(\frac{d_i}{n_i(n_i-d_i)}\right). \tag{16}$$

These variance estimates may extend above and below zero, which violates survival curve assumptions. A common fix [38] which provides confidence intervals for NA estimate as well is to estimate the variance of the log-log transformed estimate:

$$\mathrm{Var}\left[g[\widehat{S}_{KM}(t)]\right]=\left[\frac{1}{\log[\widehat{S}_{KM}(t)]}\right]^2\prod_{t_i\leq t}\left(\frac{d_i}{n_i(n_i-d_i)}\right), \tag{17}$$

where $g(u)=\log[-\log[u]]$. Transforming back with $g^{-1}(u)=\exp[-\exp[u]]$ gives the KM C.I. in Table III.

TABLE IV
SIMPLE SURVIVAL TYPE DISTRIBUTIONS

Function	$S(t)$	$H(t)$	$h(t)$
Exponential	$e^{-\lambda t}$	λt	λ
Weibull	$e^{-(\lambda t)^\alpha}$	$(\lambda t)^\alpha$	$\lambda\alpha(\lambda t)^{\alpha-1}$
Log-Logistic	$\frac{1}{1+(\lambda t)^\alpha}$	$\log\left[1+(\lambda t)^\alpha\right]$	$\frac{\lambda\alpha(\lambda t)^{\alpha-1}}{1+(\lambda t)^\alpha}$
Log-Normal[a]	$1-\Phi\left(Z(t)\right)$	$\log\left[\frac{1}{1-\Phi\left(Z(t)\right)}\right]$	$\frac{\phi(Z(t))}{\sigma t\left[1-\Phi\left(Z(t)\right)\right]}$

[a] $Z(t)=\frac{ln(t)-\mu}{\sigma}$

D. Playtime Survival for Game Data

The Exponential, Weibull, Log-Normal and Log-logistic distributions listed in Table IV are common parametric models for survival data [36]. In Figure 6 we have fitted these four models using ML to Hipster Sheep playtimes. We observe three models with significant model deviations. The exponential distribution overestimates early survival and underestimates late survival. The Log-Normal and Log-Logistic distributions fit short playtimes but have significantly longer tails than observed in practice. The Weibull distribution appears to have least model deviation, corroborating the finding that it provides good approximations to multiple games [25].

Figure 6 demonstrates why the nonparametric Kaplan-Meier and Nelson-Aalen estimates are popular. Parametric models are more powerful wherever they describe the data, but when they do not the results are incorrect. If the parametric form is correctly specified, less data is required to approximate the population distribution and they may be used to extrapolate outside the sample. Nonparametric methods are however robust to model deviations, in other words they are often the safe choice when the distribution is unknown. Even in limited data sets they are often sufficient to describe quantities of interest [23]. The confidence intervals provided for both types of estimates are informative in data constrained industry applications. Since user acquisition and software development costs money, a manager might request a user test with a limited data set or require statistical significance before committing to a major change, which makes confidence intervals useful.

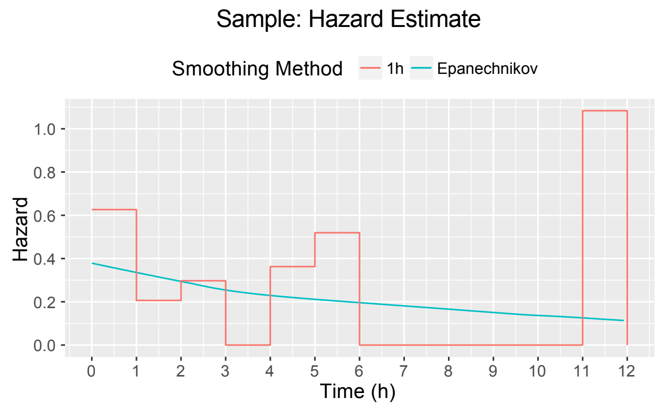

Fig. 7. Smoothing the 10 player sample with 1 hour piecewise exponential rates contrasted to Epanechnikov kernel with $b = 9.6$. These bins are too small for this sample, and a higher degree of smoothing appears more informative.

TABLE V
SAMPLE: KM-BASED AREA COMPUTATION

Time t_i	At risk n_i	Length $t_i - t_{i-}$	Survival S_{i-1}	Add Area A_i	Tail Area B_i
0	NA	NA	NA	NA	3.28
0.00	10	0.00	1.00	0.00	3.27
0.03	9	0.03	0.90	0.03	3.25
0.18	8	0.15	0.80	0.12	3.13
0.38	7	0.20	0.70	0.14	2.99
1.85	5	1.47	0.60	0.88	2.11
2.36	4	0.51	0.48	0.25	1.86
4.76	3	2.39	0.36	0.86	1.00
5.93	2	1.17	0.24	0.28	0.72
11.92	1	6.00	0.12	0.72	NA

VI. PLAYTIME METRICS

In this section three important metrics are described to benchmark the quality of a game. The metrics are motivated by the survival curve: the hazard, the mean playtime and the median playtime. These metrics are simple, easy to measure and it is possible to assess their reliability with confidence intervals.

A. Hazard as a Metric

In reliability engineering, the failure rate is a key measure of product reliability [36]: it provides a profile of how reliability evolves over time. Products may experience early failures due to defective units, the rate may then stabilize to a constant for the period of 'useful life' and go up in the 'worn-out' period. The churn rate provides a similar funnel type visual estimate for games and can be investigated in terms of the early, middle and late-game hazards. In general, the hazard is an informative time-dependent metric for the risk of the event in occurring.

In free-to-play games, it is often observed that the failure rate is very high during initial sessions, and stabilizes or steadily continues to decline as most dedicated players remain [12]. In pay-to-play games with campaigns, one may observe playtimes that are more clustered [21]. This analysis could prove useful for game-design as well [34]. In terms of game progression, good level design should have an approximately uniform churn: unexpected increases in the churn rate signify flaws and level specific decreases suggest underutilized improvements.

A major problem with small sample hazard estimation is that the interpretation could depend on the method chosen. This is illustrated with the 10 player sample in Figure 7. The piecewise exponential method has a constant hazard, or exponential distribution, within given pieces (bins) of the domain. With 1 hour bins, there are $d_{1h} = 4$ churns within the first hour with total time at risk $T_{1h} = 0.38 + 1_{+} + 0.18 + 0.00 + 1_{+} + 1_{+} + 0.79_{+} + 1_{+} + 1_{+} + 0.03 = 6.38$ h, making the first 1 hour rate λ_{1h} = 4/6.38 = 0.63 churns/h. In bins with no churns the rate is 0.00, and in the last bin there is 1 churn with a single player at risk for time $T_{12h} = 0.92$ h, implying λ_{12h} = 1.09 churns/h.

B. Mean Playtime as a Metric

The hazard function is not a single measure, but a set of measures, one for every point in time. Often a single measure is required to benchmark the game. If we assume monetization is roughly proportional to retention it is desirable to use the expected playtime as a singular metric to predict the total user value. For a playtime distribution the mean playtime is defined:

$$\mathbb{E}\left[T\right] = \int_{t=0}^{\infty} t f\left(t\right) dt. \quad (18)$$

The mean playtime has a surprising connection to the playtime survival: it is the area under the curve (AUC) [36]:

$$\mathbb{E}\left[T\right] = \int_{t=0}^{\infty} S\left(t\right) dt. \quad (19)$$

Therefore, to compare two survival curves using a single metric one may compare the area underneath each. This is quite remarkable, since we then have a singular statistic quantifying the goodness of a game. The comparison is well-defined even in cases where the survival curves cross and the relative ranking is time-dependent. The metric quantifies how much better in additional mean playtime one survival curve is. Visually the shape of the survival curve describes where the additional playtime has been accumulated from: one may have achieved it by decreasing initial churn or increasing long-term retention. In a trade-off situation, the survival curve with a larger area is better relative to this metric.

The method of deriving an estimate for the mean through the area under the survival curve is beneficial because it works with censored observations. Simply ignoring the censored observations would lead to a downward bias in the estimate. Furthermore, confidence intervals for the mean can be derived utilizing the area. In Table V. we have computed intervals $t_i - t_{i-1}$ between churn events and highlighted how much each interval adds to the total area $A_i = S_{i-1}\left(t_i - t_{i-1}\right)$. The tail area $B_i = \sum_{k=i+1}^{m} A_k$, denotes the area remaining in the tail after all areas have been accounted up to i. The total area $A = B_0 = \sum_{i=1}^{m} A_i$, which is the expected playtime, can be computed as 3.28 hours in this example.

95% confidence intervals using the normal approximation are obtained with $A \pm 1.96\sqrt{\text{Var}\left[A\right]}$, in this case 3.28 ± 2.47 hours. The variance estimate for A can be derived [35]:

$$\text{Var}\left[A\right] = \sum_{i=1}^{m} \frac{B_i^2}{n_i\left(n_i - 1\right)}. \quad (20)$$

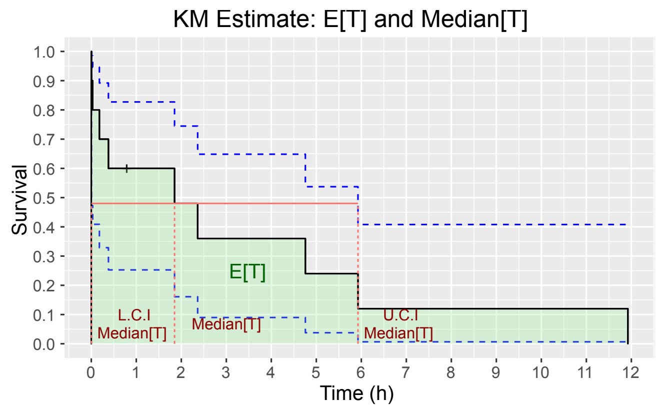

Fig. 8. Reading E[T] and Median[T] from the KM survival curve. E[T] is the area under the curve (AUC) whereas the curve drops below 0.5 at Median[T] with confidence intervals necessitated by the KM confidence intervals in blue.

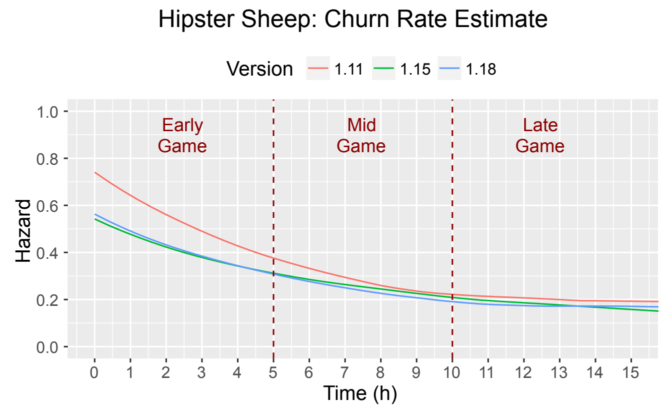

Fig. 9. The version hazard estimates computed with Epanechnikov-kernels using a high degree of smoothing $b = 9.6$ and left continuity correction. First version seems uniformly worse, whereas other two appear indistinguishable.

C. Median Playtime as a Metric

The mean playtime is quite informative in many cases, but it may not quantify typical player experience due to many early failures or the presence of a long tail. The median on the other hand attempts to quantify the 'typical' playtime. It is defined as the point to which half of the players survived:

$$\text{Median}\,[T] = \min\{t \mid S(t) \leq 0.5\}. \tag{21}$$

In general, one can define arbitrary quantiles for the survival curve. Specifically, the quantiles are defined:

$$\text{p'th quantile}\,[T] = \min\{t \mid S(t) \leq 1-p\}. \tag{22}$$

To compare two survival curves, one may compare the points at which half of players are lost. The game with the greater time to lose half of the players is then said to be better, relative to the median. This benchmark may be extended by creating quantile measures, such as a sequence of times $\{t_{10\%}, t_{20\%}, \ldots, t_{100\%}\}$ at which $10\%, 20\%, \ldots, 100\%$ of players are lost for each game. These measures give an unequivocal benchmark of short term and long term retention.

The median playtime can be read from the survival curve by finding the earliest t at which the survival drops to equal to or below 0.5. Furthermore, confidence intervals for the median can also be directly read from the pointwise KM-estimate confidence intervals: one draws a vertical line from the median and reads the left (lower) and right (upper) T value at x-axis where the line meets the KM C.I. Specifically, we seek lowest and highest value of t such that the following inequality with the log-log transform $g(u) = \log[-\log[u]]$ is satisfied [37]:

$$-z_{\alpha/2} \leq \frac{g\left(\widehat{S}(t)\right) - g(0.5)}{\sqrt{\text{Var}\left[g\left(\widehat{S}(t)\right)\right]}} \leq z_{\alpha/2}. \tag{23}$$

where $\text{Var}\left[g\left(\widehat{S}(t)\right)\right]$ was estimated previously to obtain the log-log transformed KM confidence intervals. The normal approximation based value is $z_{\alpha/2} = 1.96$ for 95% C.I. The p'th quantile confidence interval may be estimated by substituting $g(0.5) \to g(p)$ in the inequality. In this case, we obtain a highly uncertain estimate $1.85\,[0.00 \leftrightarrow 5.93]$ (h).

TABLE VI
HIPSTER SHEEP: PLAYTIME METRICS

Version	Mean	CI.l. 95%	CI.u. 95%	Median	CI.l. 95%	CI.u. 95%
1.11	1.55	1.37	1.73	0.60	0.55	0.68
1.15	2.21	2.00	2.42	0.77	0.66	0.87
1.18	2.41	2.18	2.64	0.77	0.66	0.86

D. Playtime Metrics for Game Data

The churn rate provides a very useful time-dependent metric of game quality. As explained previously, it can be used as a funnel to quantify strong and weak points of the game. In games with long-term consumption patterns, a simple hazard enables reliable player lifetime forecasts.

In Figure 9. we used Epanechnikov-kernels to obtain a hazard estimate for the three versions in the data set. Since the playtime is terminated by a churn event, this is a smooth estimate of the churn rate. We see that version 1.18 churn is quite high initially at about 0.6 churns/h, and halves to 0.3 churns/h during the first 4 hours, designated as early gameplay. A steady decline continues as most dedicated players remain in the game and the rate appears to reach near constant 0.2 churns/h from 10 hours onwards.

For the singular metrics that summarize the survival curve, Table VI. computes the mean and the median with confidence intervals for the three game versions in the data set. One should note that the mean confidence interval of 1.11 contrasted to either 1.15 or 1.18 does not overlap, which implies the difference is statistically significant. However, between 1.15 and 1.18 either confidence interval is wide enough to contain the other mean estimate. The quantile metrics, including the median, may be read with confidence intervals from the KM estimates for the versions in Figure 10.

The difference between the mean and median as metrics is clearly visible; whereas the players in 1.18 typically quit after 0.77 hours, the expected playtime extracted out of these players is three times larger at 2.41 hours. The presence of both fickle and dedicated players produces effects which make both metrics informative from different managerial perspectives.

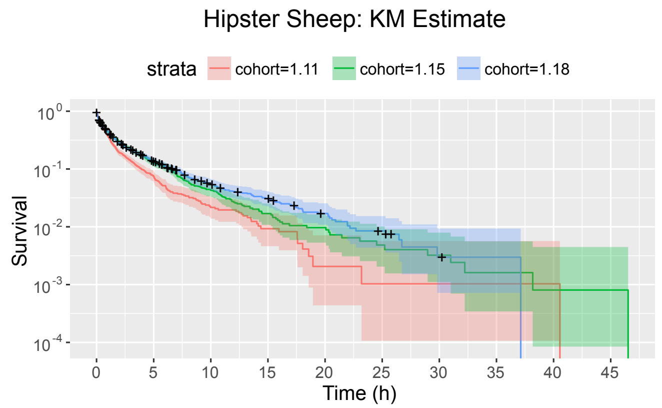

Fig. 10. Kaplan-Meier estimator for the entire population in Hipster Sheep divided into cohorts by three versions. There is some censoring in the most recent version. Based on the plot, one might think that the early version 1.11 is the worst, but the improvement from 1.15 to 1.18 is harder to deduce.

VII. Playtime Comparison

A. Comparing Cohort Survival

Given two survival curves, a test of their difference is often asked for. Statistical tests can be used for this purpose. These tests assume that the samples are identically distributed under a null hypothesis, and we obtain evidence which may reject this conclusion within a given degree of confidence. For example, we can have two game versions and survival curves produced by two cohorts consisting of players for each game version. We then assume the changes had no effect, that the survival is equal, and the evidence given by players provides a test which may reject this assumption, leading us to conclude that indeed the changes affected the game.

While several statistical tests exist, it is useful to be able to compare two survival curves in their entirety. Instead of comparing the means $\mathrm{AUC}\left[S_1\left(t\right)\right] = \mathrm{AUC}\left[S_2\left(t\right)\right]$ or the possibility that one is strictly better $S_1\left(t\right) < S_2(t)$, we present the log-rank test [23] which tests the assumption that $S_1(t) = S_2(t)$ under censored observations and allows one to use the survival curve to ascertain the difference. For simplicity we call one of the cohorts a control cohort and the other a test cohort.

Suppose that at time t_i there are n_{0i} players with d_{0i} churning in the control cohort and n_{1i} players with d_{1i} churning in the test cohort. Denote $n_i = n_{0i} + n_{1i}$ total players and $d_i = d_{0i} + d_{1i}$ total churning. The log-rank test is based on the observation that if the null hypothesis was true, the groups were equal, then given n_{0i}, n_{1i}, and d_i, the number d_{0i} is an observation of a hypergeometric random variable D_{0i}:

$$\mathbb{P}\left(D_{0i} = d_{0i} \mid n_{0i}, n_{1i}, d_i\right) = \frac{\binom{n_{0i}}{d_{0i}}\binom{n_{1i}}{d_{1i}}}{\binom{n_i}{d_i}}. \tag{24}$$

The mean and the variance of this distribution are [35]:

$$E\left[D_{0i}\right] = \frac{n_{0i} d_i}{n_i} \qquad \mathrm{Var}\left[D_{0i}\right] = \frac{n_{0i} n_{1i} d_i \left(n_i - d_i\right)}{n_i^2 (n_i - 1)}. \tag{25}$$

TABLE VII
Statistical Test of Survival Equivalence

Test	N_{control}	N_{test}	$U^2/\mathrm{Var}\left[U\right]$	p-value
$S_{1.15}\left(t\right) = S_{1.11}\left(t\right)$	1246	970	21.4	3.74e-06
$S_{1.15}\left(t\right) = S_{1.18}\left(t\right)$	1246	1537	0.4	0.534

Using these facts, it is possible to construct a linear test statistic based on a score statistic obtained by summing the differences between observed and expected event counts [35]:

$$U = \sum_{i=1}^{m}\left(d_{0i} - E\left[D_{0i}\right]\right) \qquad \mathrm{Var}[U] = \sum_{i=1}^{m}\mathrm{Var}\left[D_{0i}\right]. \tag{26}$$

A chi-square test statistic allows one to obtain a p-value [35]:

$$\frac{U^2}{\mathrm{Var}\left[U\right]} \sim \chi_1^2. \tag{27}$$

To apply this test in a real example, in Table VII we have taken the version 1.15 as a control group and compared it to a cohort with version 1.11 and then to 1.18, which were plotted in Figure 10. The first test then answers the question whether the version 1.15 was an improvement over 1.11 and the second whether the version 1.18 further improved the game. The difference between 1.11 and 1.15 is highly significant. However, the difference between 1.15 and 1.18 is completely non-significant, implying that the visually difference in Figure 10 could be caused by sampling.

There are modifications to the log-rank test which emphasize different aspects of the survival curve: one may want to weight early or late failures more heavily. For example, the weighted test statistics $U = \sum_{i=1}^{m} w_i \left(d_{0i} - E\left[D_{0i}\right]\right)$ and $\mathrm{Var}\left[U\right] = \sum_{i=1}^{m} w_i^2 \mathrm{Var}\left[D_{0i}\right]$ are commonly used with the weight $w_i = m\widehat{S}\left(t_i\right)^{\rho}$ [37]. Setting $\rho = 1$ one obtains the Prentice or Peto-Peto modification of the Gehan-Wilcoxon test which places more emphasis on earlier survival differences. In our case, this modification resulted in p-values 0.004 and 0.76 which have the same interpretation.

B. Stratification

The cohorts may not always be directly comparable. For example, user acquisitions may be conducted with different marketing campaigns or in different countries. Therefore, differences between two versions might really reflect a different underlying composition of players and not changes in behavior due to the versions themselves.

To correct for such effects, one needs to adjust for the covariate which is suspected to be an alternate cause for the effects. This is equivalent to testing the null hypothesis $S_{1j}\left(t\right) = S_{2j}(t)$ across groups $j = 1, \ldots, G$. The test is based on computing score statistic U_g and variance $\mathrm{Var}[U]$ for each group separately and using the test [37]:

$$\frac{\left(\sum_{g=1}^{G} U_g\right)^2}{\sum_{g=1}^{G} \mathrm{Var}\left[U_g\right]} \sim \chi_1^2\,. \tag{28}$$

In our case adjusting for the country of origin, we obtain p-values 7.88e-06 and 0.608 which again does not change the previous interpretation.

TABLE VIII
APPLIED RESEARCH CHEATSHEET

Problem	R function [37]	R library
Fit a nonparametric survival model?	survfit	survival
Fit a parametric survival model?	survreg	survival
Fit a nonparametric hazard?	muhaz/pehaz	muhaz
Compute the mean or median?	print(print.rmean=T/F)	survival
Compute the log-rank test?	survdiff	survival

VIII. CONCLUSION

The purpose of this study was to investigate whether survival analysis methods contribute to the analysis of player retention and churn. The findings suggests that survival analysis motivated functions, metrics and comparisons provide multiple tools to utilize for retention measurement in game development. Retention as user engagement can be quantified with several duration variables such as playtimes, session lengths, subscription times, and even game progression. In this study, we focused on total playtime in the interest of it providing a straightforward measure of total engagement.

From a practical point of view, the survival curve is a visual funnel type metric and the associated statistics provide a set of new metrics to game analytics. Since the metrics were derived from the survival curve, they work for censored data and the uncertainty can be estimated with confidence intervals. The hazard describes the continuous churn rate and offers a clear time-dependent game quality interpretation. The mean, the median and the quantile estimates as playtime metrics can be used to aggregate the survival data to a single statistic to measure different aspects of the playtime phenomena. For survival curve comparison, the log-rank is a test of the null hypothesis that the survival curves are equal. The test may be extended to stratify over covariates and compare multiple survival curves. This method enables scientific AB-testing of game version quality relative to a duration metric such as the total playtime, and can be used to verify that a visually observed difference is statistically significant.

These methods contribute towards scientific data analysis by presenting new metrics to game analytics which are also able to deal with censoring and utilize statistical significance tests. The reader may take advantage of Table VIII to use the methods for applications. It lists the methods we have presented and the R software functions implementing them.

Naturally, we acknowledge that this research has some limitations. These methods have an implicit challenge associated with churn uncertainty, which is scarcely addressed within the fields of survival analysis or game analytics. When a measurement depends on user churn, impromptu rules currently in use could lead to bias. Future research that assesses and possibly mitigates this is called for. While our data is from one game only and focused on total playtime, the methods are directly transferable to any duration data. Further studies could be done to investigate the kind of phenomena present in different types of games. We think that survival analysis has a large potential to contribute to scientific game analytics and anticipate further research on this topic.

ACKNOWLEDGMENT

We thank Tribeflame Ltd. for their continuing participation and Anne-Maarit Majanoja for helpful feedback.

REFERENCES

[1] M. S. El-Nasr, A. Drachen, and A. Canossa, Eds., *Game Analytics, Maximizing the Value of Player Data*. Springer, 2013.
[2] E. B. Seufert, "Freemium metrics," in *Freemium Economics*. Boston: Morgan Kaufmann, 2014, pp. 83–113.
[3] B. G. Weber, M. Mateas, and A. Jhala, "Using data mining to model player experience," in *Proc. FDG Evaluating Player Experience in Games Workshop*. ACM, 2011.
[4] B. G. Weber, M. John, M. Mateas, and A. Jhala, "Modeling player retention in madden NFL 11," in *Proc. Innovative Applications of Artificial Intelligence Conf.*, D. G. Shapiro and M. P. J. Fromherz, Eds. AAAI Press, 2011.
[5] B. E. Harrison and D. L. Roberts, "When players quit (playing Scrabble)," in *Proc. AAAI Artificial Intelligence and Interactive Digital Entertainment Conf.*, M. Riedl and G. Sukthankar, Eds. The AAAI Press, 2012.
[6] B. Harrison and D. L. Roberts, "Analytics-driven dynamic game adaption for player retention in Scrabble," in *Proc. IEEE Computational Intelligence in Games Conf.*, 2013, pp. 1–8.
[7] B. E. Harrison and D. L. Roberts, "Analytics-driven dynamic game adaption for player retention in a 2-dimensional adventure game," in *Proc. AAAI Artificial Intelligence and Interactive Digital Entertainment Conf.*, 2014.
[8] B. Harrison and D. L. Roberts, "An analytic and psychometric evaluation of dynamic game adaption for increasing session-level retention in casual games," *IEEE Trans. Comput. Intell. AI in Games*, vol. 7, no. 3, pp. 207–219, 2015.
[9] T. Debeauvais, C. V. Lopes, N. Yee, and N. Ducheneaut, "Retention and progression: Seven months in World of Warcraft," in *Proc. Int. Foundations of Digital Games Conf.*, M. Mateas, T. Barnes, and I. Bogost, Eds. Society for the Advancement of the Science of Digital Games, 2014.
[10] T. Debeauvais and C. V. Lopes, "Gate me if you can: The impact of gating mechanics on retention and revenues in Jelly Splash," in *Proc Int. Foundations of Digital Games Conf.*, J. P. Zagal, E. MacCallum-Stewart, and J. Togelius, Eds. Society for the Advancement of the Science of Digital Games, 2015.
[11] M. Viljanen, A. Airola, T. Pahikkala, and J. Heikkonen, "Modelling user retention in mobile games," in *Proc. IEEE Computational Intelligence and Games Conf.* IEEE, 2016, pp. 62–69.
[12] ——, "User activity decay in mobile games determined by simple differential equations?" in *Proc. IEEE Computational Intelligence and Games Conf.* IEEE, 2016, pp. 126–133.
[13] J. Kawale, A. Pal, and J. Srivastava, "Churn prediction in MMORPGs: A social influence based approach," in *Proc. Int. Computational Science and Engineering Conf.*, vol. 4, 2009, pp. 423–428.
[14] Z. Borbora, J. Srivastava, K. W. Hsu, and D. Williams, "Churn prediction in MMORPGs using player motivation theories and an ensemble approach," in *IEEE Int. Privacy, Security, Risk and Trust Conf. and IEEE Int. Social Computing Conf.*, 2011, pp. 157–164.
[15] Z. H. Borbora and J. Srivastava, "User behavior modelling approach for churn prediction in online games," in *IEEE Int. Privacy, Security, Risk and Trust Conf. and IEEE Int. Social Computing Conf.*, 2012, pp. 51–60.
[16] J. Runge, P. Gao, F. Garcin, and B. Faltings, "Churn prediction for high-value players in casual social games," in *Proc. IEEE Computational Intelligence and Games Conf.*, 2014, pp. 1–8.
[17] P. Rothenbuehler, J. Runge, F. Garcin, and B. Faltings, "Hidden Markov models for churn prediction," in *Proc. SAI Intelligent Systems Conf.*, 2015, pp. 723–730.
[18] F. Hadiji, R. Sifa, A. Drachen, C. Thurau, K. Kersting, and C. Bauckhage, "Predicting player churn in the wild," in *Proc. IEEE Computational Intelligence and Games Conf.*, 2014, pp. 1–8.
[19] M. Tamassia, W. Raffe, R. Sifa, A. Drachen, F. Zambetta, and M. Hitchens, "Predicting player churn in Destiny: A hidden Markov models approach," in *Proc. IEEE Computational Intelligence and Games Conf.* IEEE, 2016, pp. 325–332.
[20] R. Sifa, S. Srikanth, A. Drachen, C. Ojeda, and C. Bauckhage, "Predicting retention in sandbox games with tensor factorization-based representation learning," in *Proc. IEEE Computational Intelligence and Games Conf.* IEEE, 2016, pp. 142–149.

[21] T. Allart, G. Levieux, S. Natkin, and A. Guilloux, "Design influence on player retention : A method based on time varying survival analysis," in *Proc. IEEE Computational Intelligence and Games Conf.* IEEE, 2016.

[22] A. G. . Periez, A. Saas and C. Magne, "Churn prediction in mobile social games: Towards a complete assessment using survival ensembles," in *Proc. IEEE International Conf. on Data Science and Advanced Analytics (DSAA)*. IEEE, 2016.

[23] D. G. Kleinbaum and M. Klein, *Survival Analysis: A Self-Learning Text (Statistics for Biology and Health)*, 2nd ed. Springer, 2005.

[24] C. Bauckhage, K. Kersting, R. Sifa, C. Thurau, A. Drachen, and A. Canossa, "How players lose interest in playing a game: An empirical study based on distributions of total playing times," in *Proc. IEEE Computational Intelligence and Games Conf.* IEEE, 2012, pp. 139–146.

[25] R. Sifa, C. Bauckhage, and A. Drachen, "The playtime principle: Large-scale cross-games interest modeling," in *Proc. IEEE Computational Intelligence and Games Conf.* IEEE, 2014, pp. 1–8.

[26] T. Henderson and S. Bhatti, "Modelling user behaviour in networked games," in *Proc. ACM International Multimedia Conf.* New York, NY, USA: ACM, 2001, pp. 212–220.

[27] M. Kwok and G. Yeung, "Characterization of user behavior in a multi-player online game," in *Proc. ACM SIGCHI Int. Advances in Computer Entertainment Technology Conf.* New York, NY, USA: ACM, 2005, pp. 69–74.

[28] C. Chambers, W.-c. Feng, S. Sahu, and D. Saha, "Measurement-based characterization of a collection of on-line games," in *Proc. ACM SIGCOMM Internet Measurement Conf.* USENIX Association, 2005, pp. 1–1.

[29] W.-c. Feng, D. Brandt, and D. Saha, "A long-term study of a popular MMORPG," in *Proc. ACM SIGCOMM Network and System Support for Games Workshop*. New York, NY, USA: ACM, 2007, pp. 19–24.

[30] P.-Y. Tarng, K.-T. Chen, and P. Huang, "An analysis of WoW players' game hours," in *Proc. ACM SIGCOMM Network and System Support for Games Workshop*. New York, NY, USA: ACM, 2008, pp. 47–52.

[31] K.-T. Chen, P. Huang, and C.-L. Lei, "Effect of network quality on player departure behavior in online games," *IEEE Trans. Parallel Distrib. Syst.*, vol. 20, no. 5, pp. 593–606, 2009.

[32] D. Pittman and C. GauthierDickey, "A measurement study of virtual populations in massively multiplayer online games," in *Proc. ACM SIGCOMM Network and System Support for Games Workshop*. New York, NY, USA: ACM, 2007, pp. 25–30.

[33] ——, "Characterizing virtual populations in massively multiplayer online role-playing games," in *Proc. Int. Advances in Multimedia Modeling Conf.* Berlin, Heidelberg: Springer-Verlag, 2010, pp. 87–97.

[34] A. Isaksen, D. Gopstein, and A. Nealen, "Exploring game space using survival analysis," in *Proc. Int. Foundations of Digital Games Conf.*, J. P. Zagal, E. MacCallum-Stewart, and J. Togelius, Eds. Society for the Advancement of the Science of Digital Games, 2015.

[35] S. Selvin, *Survival Analysis for Epidemiologic and Medical Research (Practical Guides to Biostatistics and Epidemiology)*, 1st ed. Cambridge University Press, 2008.

[36] J. F. Lawless, *Statistical models and methods for lifetime data*, ser. Wiley series in probability and mathematical statistics. Applied probability and statistics. New-York, Chichester, Brisbane: J. Wiley, 1982.

[37] D. F. Moore, *Applied survival analysis using R*, ser. Use r! Cham: Springer, 2016.

[38] R. J. Cook and J. F. Lawless, *The statistical analysis of recurrent events*, ser. Statistics for biology and health. New York, London: Springer, 2007.

[39] F. E. Harrell, *Regression Modeling Strategies, with Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis*, second edition ed. New York: Springer, 2015.