Is High-Altitude Mountaineering Russian Roulette?

Edward K. Cheng*

Prepublication draft of February 22, 2013
(forthcoming in the *Journal of Quantitative Analysis in Sports* (2013))

1 Introduction

In their respective memoirs, mountaineers David Roberts and Ed Viesturs express a fundamental disagreement over the risks associated with climbing high-altitude (8000m) peaks. (Viesturs and Roberts, 2006, Roberts, 2005). For Roberts, the risk of dying on an expedition to an 8000m peak is effectively static, and so he likens it to Russian roulette. (Viesturs and Roberts, 2006, 308) Based on a 1997 study showing that "on any given expedition to an [8000m] peak…a climber stands a 1-in-34 chance of dying," Roberts calculates that Viesturs had cumulatively faced a 60 percent chance of dying over his thirty expeditions on 8000m peaks.[1] (Roberts, 2005, 348-49).

For Viesturs, nothing could be more "ridiculous." (Viesturs and Roberts, 2006, 308). Viesturs maintains that "most accidents and deaths on the high peaks are due to human error," and that his experience and attention to safety made his ascents far less risky than for the typical climber.(Viesturs and Roberts, 2006, 310). Emphatically rejecting Roberts's analogy of elite mountaineering to Russian roulette, Viesturs contends that with each climb, he became a "better mountaineer—smarter, faster, stronger, more efficient," making each subsequent climb safer than the previous. (Viesturs and Roberts, 2006, 311-12).

Although readers may viscerally side with Roberts or Viesturs, neither position is inherently unreasonable. On the one hand, in many walks of life, greater experience and skill

*Professor of Law, Vanderbilt University Law School; Doctoral Student, Department of Statistics, Columbia University. Thanks to Jim Albert, Andreas Buja, Andrew Gelman, Ray Huey, David Madigan, Zach Shahn, and two anonymous reviewers for helpful comments and conversations. Research was conducted under Columbia University IRB protocol IRB-AAAF3302 (exempt) and supported by a Bedayn Research Grant from the American Alpine Club.

[1] This 60 percent figure would appear to come from $1-(1-\frac{1}{34})^{30}=0.592$.

correlate with better outcomes, so why not with mountaineering? Additionally, one can imagine something akin to Darwinian selection occurring on Himalayan slopes, with risky (or perhaps less fit) climbers more likely becoming early casualties, leaving only safer and stronger climbers to attempt multiple ascents. On the other hand, perhaps the largely unavoidable risks, what some term “objective” risks—avalanches, crevasses, falling rocks, weather, etc.—simply overwhelm what little advantage greater experience and skill provide. The notion that experience and skill affect outcomes may merely be a climber’s hollow hope about controlling something that he or she fundamentally cannot.

This study sheds some statistical light on the Roberts-Viesturs debate and assesses whether experience is associated with lower fatality risk on 8000m peaks. Using a comprehensive dataset of high-altitude climbing expeditions in the Nepalese Himalaya, the study will estimate the conditional death risk for a member of an 8000m expedition given that climber’s number of previous expeditions. Conditional death risks are after all the core of the Roberts-Viesturs controversy. Roberts contends that the risk is constant (unconditional); Viesturs suggests that the risk of death should decrease with each ensuing climb.

After estimating these conditional probabilities, I test to see if they are consistent with a constant failure rate model (Roberts) or a decreasing failure rate model (Viesturs). To perform the hypothesis testing, I propose using Total Time on Test (TTT) plots, a graphical technique derived from the reliability engineering literature. In particular, I show that this established graphical technique can be appropriately adapted to the discrete-data, censored-data context, and that graphical inference can be done either by applying some insightful ideas suggested by Buja, Cook, Hofmann, Lawrence, Lee, Swayne, and Wickham (2009), or by developing a test statistic using the TTT plots.

Finally, I augment these basic results with three supplemental analyses. First, I offer some power calculations to justify the TTT method. Second, to control for possible confounding factors, such as year and mountain, I use a discrete-time survival analysis to model baseline hazard probabilities, which are then tested for constant or decreasing rates. Finally, to provide a useful juxtaposition, I examine the probability of ascent on 8000m peaks, asking whether previous 8000m experience increases the probability of reaching the summit.

2 Death Risks in the Mountaineering Literature

Interest in the basic fatality rates associated with high-altitude mountaineering has existed for quite some time.(Windsor, Firth, Grocott, Rodway, and Montgomery, 2009, Firth, Zheng, Windsor, Sutherland, Imray, Moore, Semple, Roach, and Salisbury, 2008, Salisbury and Hawley, 2007, Burtscher, Philadelphy, Nachbauer, and Likar, 1995, Pollard and Clarke, 1988). Recently, some researchers, most notably Raymond Huey and Xavier Eguskitza, have begun testing whether factors such as elevation, gender, age, weather, or the use of bottled oxygen are associated with success or death on high-altitude peaks. For example, Huey (2001) reports an association among altitude, supplemental oxygen use, and death rate, and Huey and Eguskitza (2001) links success and death rates with elevation. (See also Huey, Salisbury, Wang, and Mao, 2007, Fontanarosa, Huey, and Eguskitza, 2000).

Two studies thus far have attempted to ascertain the effects of experience and skill on death risk. Westhoff, Koepsell, and Littell (2012) closely tracks this study, using the same dataset, but using logistic regression analysis rather than the TTT plots and graphical inference techniques advocated here. They find that experience has no effect on mortality rates. Boyce and Bischak (2010) also analyzes the effect of technological improvements and climber experience on an expedition's probability of success or misfortune. They find that "experience level has a positive effect upon ascents, but [that there are] no robust effects on . . . other outcomes [including death]."(Boyce and Bischak, 2010). Notably, Boyce and Bischak (2010) focuses on risks and probabilities at the expedition level. This expedition-level focus limits the study's usefulness for our question, which revolves around the individual.

3 Data

The dataset for this study was extracted from the Himalayan Database compiled by Elizabeth Hawley with Richard Salisbury and published by the American Alpine Club. (Salisbury, 2003). Hawley, a longtime American journalist based in Kathmandu, is something of a legendary figure in Himalayan mountaineering. She has kept records of climbs in the Nepalese Himalaya since the 1960s, and has effectively become the final arbiter of climbing claims in the region. (Jolly, 2010). Her personal archive, "supplemented by information gathered from books, alpine journals and correspondence with Himalayan climbers," forms the basis of the Himalayan Database.

For each climb, the dataset records among other things whether a successful ascent was made, whether the climber died, and the maximum altitude achieved. Of the 26,708 attempts recorded, there were 9221 ascents (34.5%) and 510 deaths (1.9%). The dataset also contains additional information including the year, peak, climber's name, sex, date of birth, nationality, and whether the climber was a leader or deputy of an expedition. This study used all climbs from 1905 to 2009. In accord with mountaineering's rise from fringe endeavor to an increasingly mainstream sport, the vast majority of climbs occur after 1990, with the median climb year being 1998.

This study used all climbers and attempts recorded in the Himalayan database with the following exceptions. First, albeit unfortunate, Sherpas were omitted because of identifiability problems. Because Sherpas often have identical names, and because birthdate information was often unavailable, there was no way to separate unique individuals with certainty. And since the study critically hinges on the death rate conditional on the number of previous climbs, the safer course was to exclude this data. Second, only climbers who went beyond base camp were considered in the study.

Of the remaining data, there were 14,077 climbers, of which 91% were male. The climbers came from at least 94 different countries, of which the most represented were the United States, Japan, Spain, United Kingdom, South Korea, France, Italy and Germany. The mean age of an mountaineer on any given attempt was about 37 years old. The distribution of ages is seen in Figure 1.

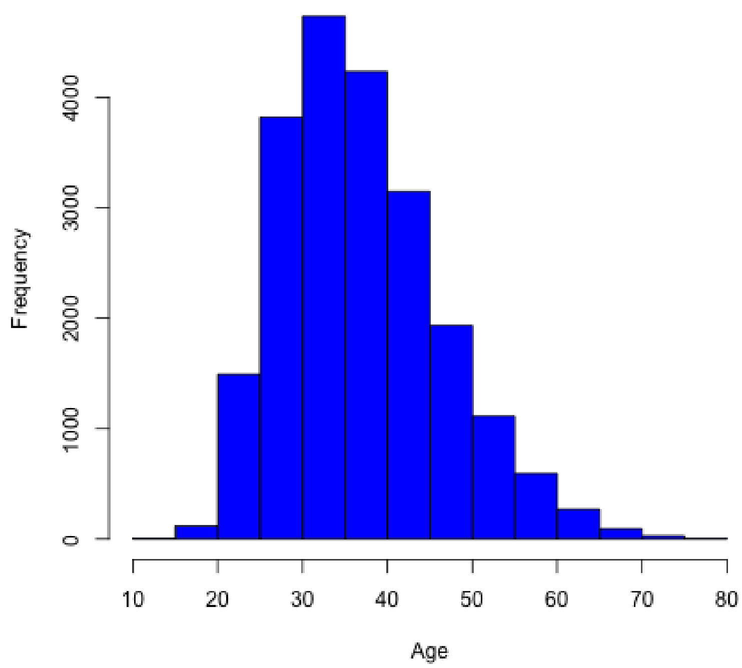

Figure 1: Histogram of Mountaineer Ages

The study includes all eight 8000m peaks covered by the Himalayan database: Everest

(8848m), Kangchenjunga (8586m), Lhotse (8516m), Makalu (8485m), Cho Oyu (8201m), Dhaulagiri (8167m), Manaslu (8163m), and Annapurna (8091m). The other six 8000m peaks: K2 (8611m), Nanga Parbat (8126m), Gasherbrum I (8080m), Broad Peak (8051m), Gasherbrum II (8034m), and Shishapangma (8027m) are located outside of Nepal and not included in the database. Notably, while Everest is the highest (and most popular) peak, it is neither the most dangerous nor the most difficult. Annapurna is generally regarded as the most perilous, and K2 the most challenging.

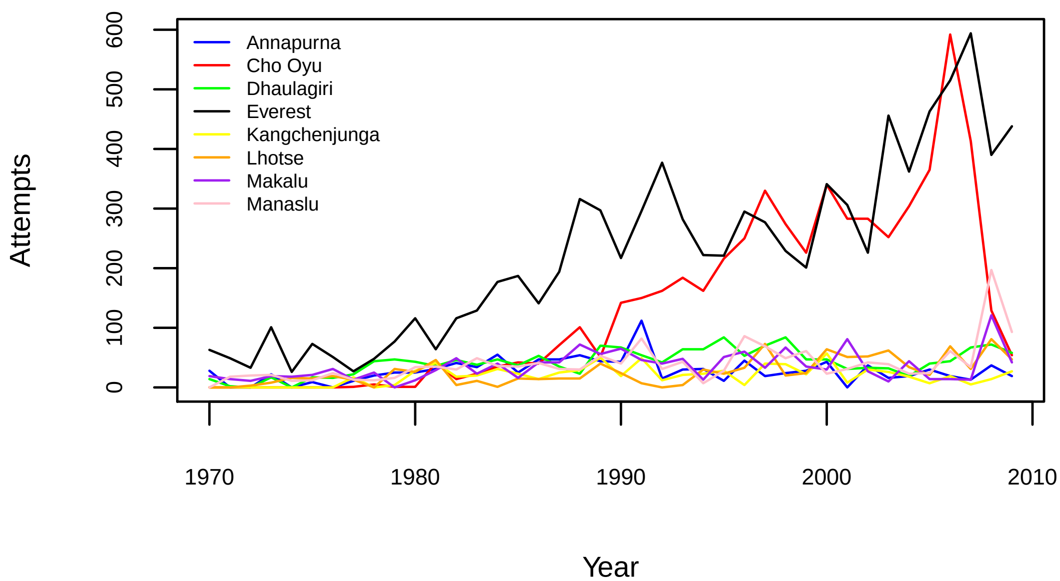

Figure 2: Attempts on Nepalese 8000m Peaks (1970-2009)

As might be expected, most climbs in the dataset represent a climber's first attempt at a Nepalese Himalayan peak. High altitude mountaineering in the Himalayas requires a significant expenditure of time (two to three months) and money. For most, it is the adventure of a lifetime. Multiple attempts are frequently only found among elite mountaineers or professional guides. Due to these sample size considerations, the study only looked at the first ten attempts for any given climber.

4 Statistical Methods

4.1 Failure Rates and Total Time on Test (TTT) Transforms

How failure rates evolve over time is a matter long studied in the reliability engineering literature, which attempts to predict the lifespan of components or products. Some products like light bulbs have a "burn-in" period, in which failure rates are relatively high at first, but decrease after those with manufacturing defects have been culled. And most products have increasing failure rates toward the end of their lifetimes as parts begin to wear out.

To study such failure rates, Barlow and Campo (1975) propose using the Total Time on Test transform. If $F(t)$ is the distribution function for F, then the Total Time on Test (TTT) Transform is defined as:

$$H_F^{-1}(v) = \int_0^{F^{-1}(v)} (1 - F(u))du \qquad (0 \leq v \leq 1)$$

and the *scaled* TTT transform is defined as:

$$\varphi_F(v) = \frac{H_F^{-1}(v)}{H_F^{-1}(1)} \qquad (0 \leq v \leq 1)$$

(Barlow and Campo, 1975, Hoyland and Rausand, 1994)

Although not immediately apparent from this construction, the key result is that the scaled TTT transform of any exponential distribution is the uniform distribution, which plots as a 45° line from the origin to (1,1). Since the exponential distribution is the only distribution with constant failure rate, we can thus use plots of scaled TTT transforms to assess whether a given distribution has constant, increasing, or decreasing failure rate. Distributions with increasing failure rates have scaled TTT transforms that are concave; distributions with decreasing failure rates have ones that are convex. (Hoyland and Rausand, 1994).

4.2 Empirical TTT Plots

To study the failure rate of an observed (but non-censored) sample, we estimate $F(t)$ in the TTT transform using the empirical distribution function $\mathbb{F}_n(t)$, and then compare the plot of the resulting scaled TTT transform with the 45° line to determine increasing or decreasing failure rates. The empirical scaled TTT transform is thus:

$$\frac{\mathbb{H}_n^{-1}(v)}{\mathbb{H}_n^{-1}(1)} = \frac{\int_0^{\mathbb{F}_n^{-1}(v)} (1 - \mathbb{F}_n(u))du}{\int_0^{\mathbb{F}_n^{-1}(1)} (1 - \mathbb{F}_n(u))du} \qquad (v = \frac{i}{n}, i = 1, 2, \ldots, n)$$

which can be shown to be a function of the so-called total time on test:

$$\frac{\mathbb{H}_n^{-1}(v)}{\mathbb{H}_n^{-1}(1)} = \frac{\mathscr{T}(T_{(i)})}{\mathscr{T}(T_{(n)})}$$

where $T_{(1)}, T_{(2)}, \ldots T_{(n)}$ denote the ordered failure times and the total time on test $\mathscr{T}(T_{(i)})$ is defined as

$$\mathscr{T}(T_{(i)}) = \sum_{j=1}^{i} T_{(j)} + (n-i)T_{(i)}.$$

(Hoyland and Rausand, 1994, 366-67)

Mountaineering data, however, is censored, since mountaineers eventually stop climbing 8000m peaks. For censored data, Hoyland and Rausand (1994) propose using a Kaplan-Meier estimator $1 - \hat{R}(t)$, rather than the empirical distribution function $\mathbb{F}_n(t)$, to construct the scaled TTT plot. In other words, we estimate $F(t)$ using $1 - \hat{R}(t)$. (See Singer and Willett, 2003, 477-480).

In particular, following Hoyland and Rausand (1994), if $T^*_{(1)}, T^*_{(2)}, \ldots, T^*_{(k)}$ are the k ordered failure times from a censored dataset, define

$$v^*_{(i)} = 1 - \hat{R}(T^*_{(i)}) \qquad (i = 1, 2, \ldots, k)$$

and

$$\hat{\mathbb{H}}^{-1}(v^*_{(i)}) = \int_0^{T^*_{(i)}} \hat{R}(u)du = \sum_{j=1}^{i-1} (T^*_{(j+1)} - T^*_{(j)})\hat{R}(T^*_{(j)}).$$

Then the scaled empirical TTT plot for censored data would consist of the points

$$\left(\frac{v^*_{(i)}}{v^*_{(k)}}, \frac{\hat{\mathbb{H}}^{-1}(v^*_{(i)})}{\hat{\mathbb{H}}^{-1}(v^*_{(k)})} \right).$$

(Hoyland and Rausand, 1994, 408-11)

4.3 Graphical Hypothesis Testing

TTT plots in practice provide a useful non-parametric method for testing constant failure rate. Indeed, as Andersen, Borgan, Gill, and Keiding (1993) suggest, the TTT plot is quite powerful — TTT plots may show a significant departure from a constant failure rate even when more traditional methods do not. (See Andersen et al., 1993, 454). In addition, empirical TTT plots are an appealing visual way to communicate failure data to a wider, non-technical audience.

The key issue is how to make inference more objective and less impressionistic in this graphical context. How much deviation from the 45° diagonal is "too much"? And surely what constitutes "too much" will change depending on the sample size, the average failure rate, and so on. In an insightful article, Buja et al. (2009) suggests using a graphical "lineup" that combines observed and simulated null data to make more rigorous inferences from visual data displays. Specifically, the actual data plot is randomly placed within an array of nineteen simulated null plots. The reader is then asked to guess the real data to produce "an inferentially valid p-value." (Buja et al., 2009, 4369-70).[2] Notably, the use of plots to compare simulated to actual data is not new, although it has predominantly found use in exploratory settings thus far. Back in the 1950s, for example, Jerzy Neyman and others famously used a photographic plate simulation of galaxy distributions to develop a model of galactic clustering. (Davies, 2008, Brillinger, 2008).

For this study, generating the simulated null data for the graphical inference procedure was a relatively straightforward task. At each time period, for each climber remaining in the active population, one draws from an appropriately estimated Bernoulli distribution to determine whether the climber dies on his/her expedition, and then from another Bernoulli distribution to determine whether the climber continues to climb or retires. Put a different way, one estimates the (constant) death probability on any climb, and the conditional probability of retiring. Then for each active climber in each time period, one flips the appropriately weighted coins to determine whether the simulated climber dies, retires, or continues.

We can further increase the objectivity of our graphical inference by forming a test statistic based on the mean negative deviation between the empirical TTT plot and the 45 degree line. We use the negative deviation so that decreasing death rates, with their convex TTT plots, will have test statistics with large positive values. On the one hand, this formal statistic necessarily sacrifices some power, since it only measures "deviation" in a specific way. On the other, it removes the subjectivity of pure visual assessment, and may improve power by quantifying differences definitively. For purposes of hypothesis testing, a null distribution for this "TTT Plot Statistic" can be estimated by simulation.

[2] Further extensions of the method might involve outsourcing the visual processing to external observers, such as those provided by a service like Amazon Mechanical Turk (mTurk). (Buja et al., 2009, 4381). Amazon mTurk is particularly useful for testing at p-values lower than 0.05, since there are limits to the number of graphs a person can simultaneously compare, and space (and boredom) limitations preclude including pages of plots in journals.

4.4 Discrete Time Survival Analysis

Although TTT plots of overall death rates on 8000m peaks are informative, one may have concerns about confounders. For example, to the extent that Mount Everest is by far the most attempted peak and that many of these attempts are very modern, one may wonder if Everest skews the results. (The direction of the distortion is not exactly clear. While modern equipment should promote greater safety, it also enables less experienced and less accomplished climbers to attempt otherwise foreclosed 8000m peaks.)

A discrete-time survival analysis model helps control for some of these possible confounders. As noted in Singer and Willett (2003), one basic model looks like:

$$cloglog(\hat{h}(t_k)) = \sum_i \alpha_i D_i + \sum_j \beta_j X_j$$

where $cloglog(x)$ is the complementary log-log link function (i.e., $\log(-\log(1-x))$), $\hat{h}(t_k)$ is the observed hazard function at a discrete time t_k, and X_j are the explanatory variables. The D_i are indicator variables for each discrete time period under consideration (i.e., $D_i = \mathbb{1}\{k = i\}$), so that the expression $\sum_i \alpha_i D_i$ represents an estimated baseline hazard function.

Having estimated the baseline hazard function, we can once again perform a TTT plot analysis to determine whether it is indeed constant, or whether it is decreasing (or increasing). Generating the simulated null data in this context requires a modest extension of the generation methods discussed previously. Since the null distribution assumes a constant failure rate, we estimate the simplified model:

$$cloglog(\hat{h}(t_k)) = \alpha_o + \sum_j \beta_j X_j$$

This model then provides the probability of death on each climb for each of the simulated climbers. As before, after each simulated climb, draws are made from the corresponding Bernoulli distributions to determine whether the simulated climber dies, retires, or continues.

5 Results

5.1 Basic Hazard (Death) Rates

Figure 3 shows the hazard probabilities for a mountaineer on a 8000m Nepalese peak given the climber's experience. Previous experience in the Nepalese Himalaya obviously is an inexact measure of experience and skill. Climbers may spend time on the other (non-Nepalese) 8000m peaks, may focus on lower altitude (but no less dangerous or technical)

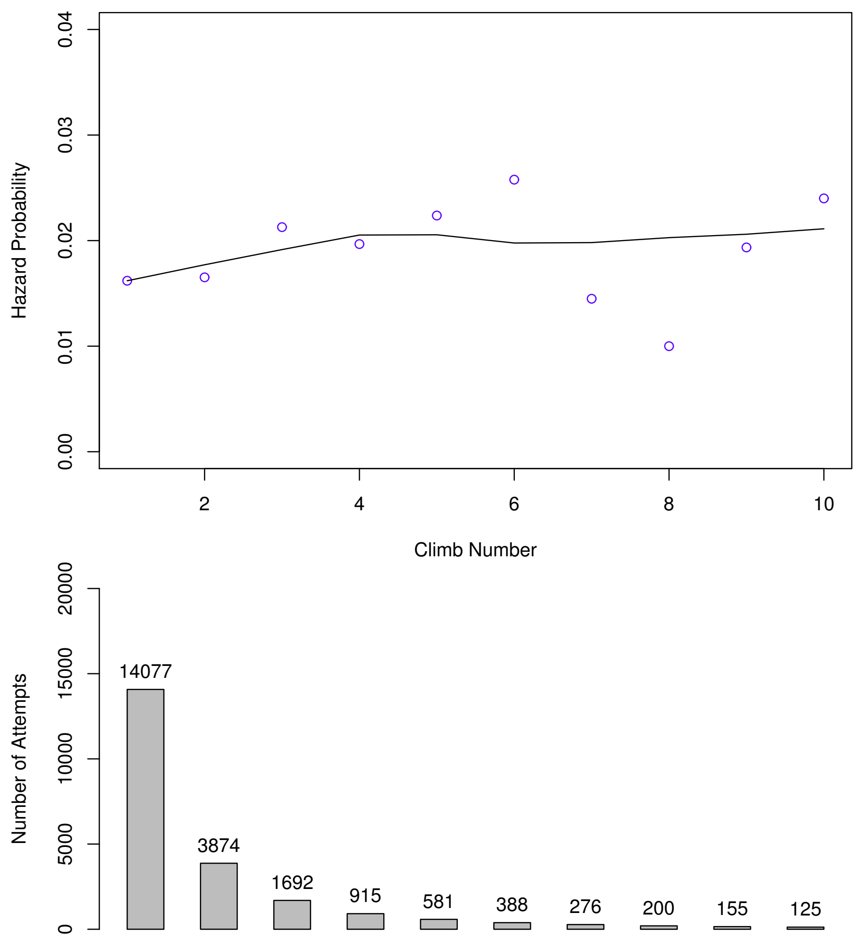

Figure 3: Conditional Hazard Probability (top) and Sample Size (bottom) by Experience

peaks, etc. Nevertheless, the previous number of climbs provides a proxy within the confines of the dataset. As Figure 3 shows, death rates typically fall within the 2% range, with the variance increasing with the number of climbs. As the sample size chart in Figure 3 suggests, this increasing variance is explained by the small number of climbers who make a large number of attempts.

The scaled empirical TTT plot in Figure 4 helps assess whether the death rate in Figure 3 is constant. As seen in Figure 4, the points seem to align closely with the 45° diagonal,

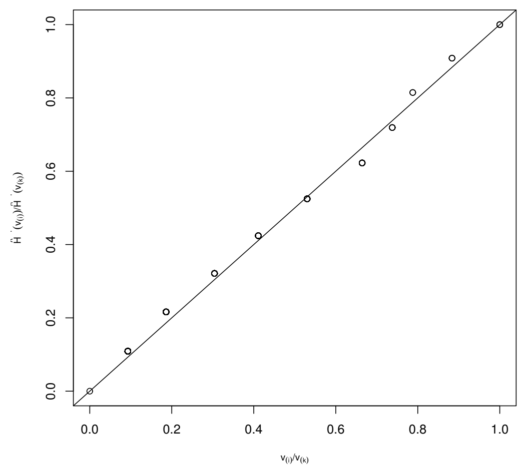

Figure 4: TTT Plot of Death Rate

suggesting that the death rate is constant regardless of the climber's experience.

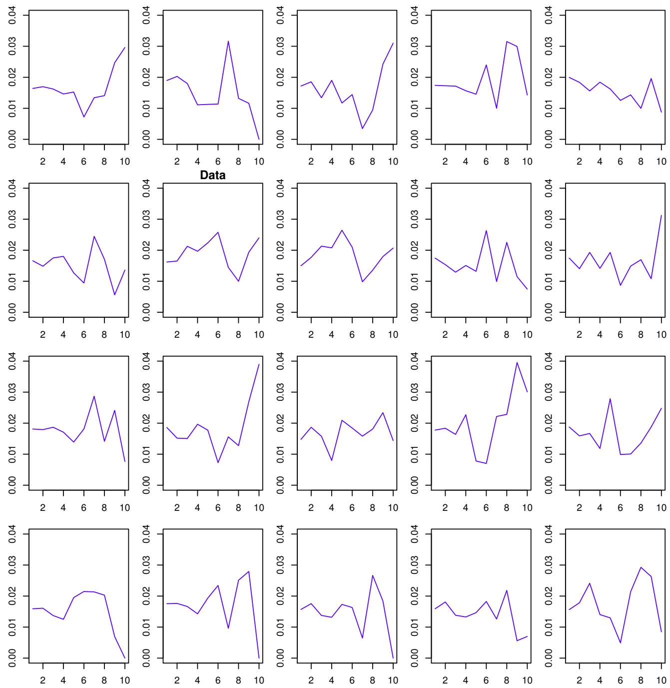

Figure 5: Line-up of Null and Observed(as indicated) Conditional Hazard Probabilities

5.2 Graphical Inference on Hazard Rates

The next step is to use the graphical inference methods previously discussed to make our visual intuitions more rigorous. From the dataset, the overall average death risk for a climber on any given climb is 0.017. In addition, we can determine the conditional withdrawal probabilities (the probabilities that a mountaineer will retire from climbing after attempting a given number of Nepalese 8000m peaks). With these two estimates in hand, we can then simulate and plot null datasets under the assumption that the death risk remains constant. Being able to detect the observed data within a set of 19 null plots would correspond to rejecting the null hypothesis at a 0.05 level.

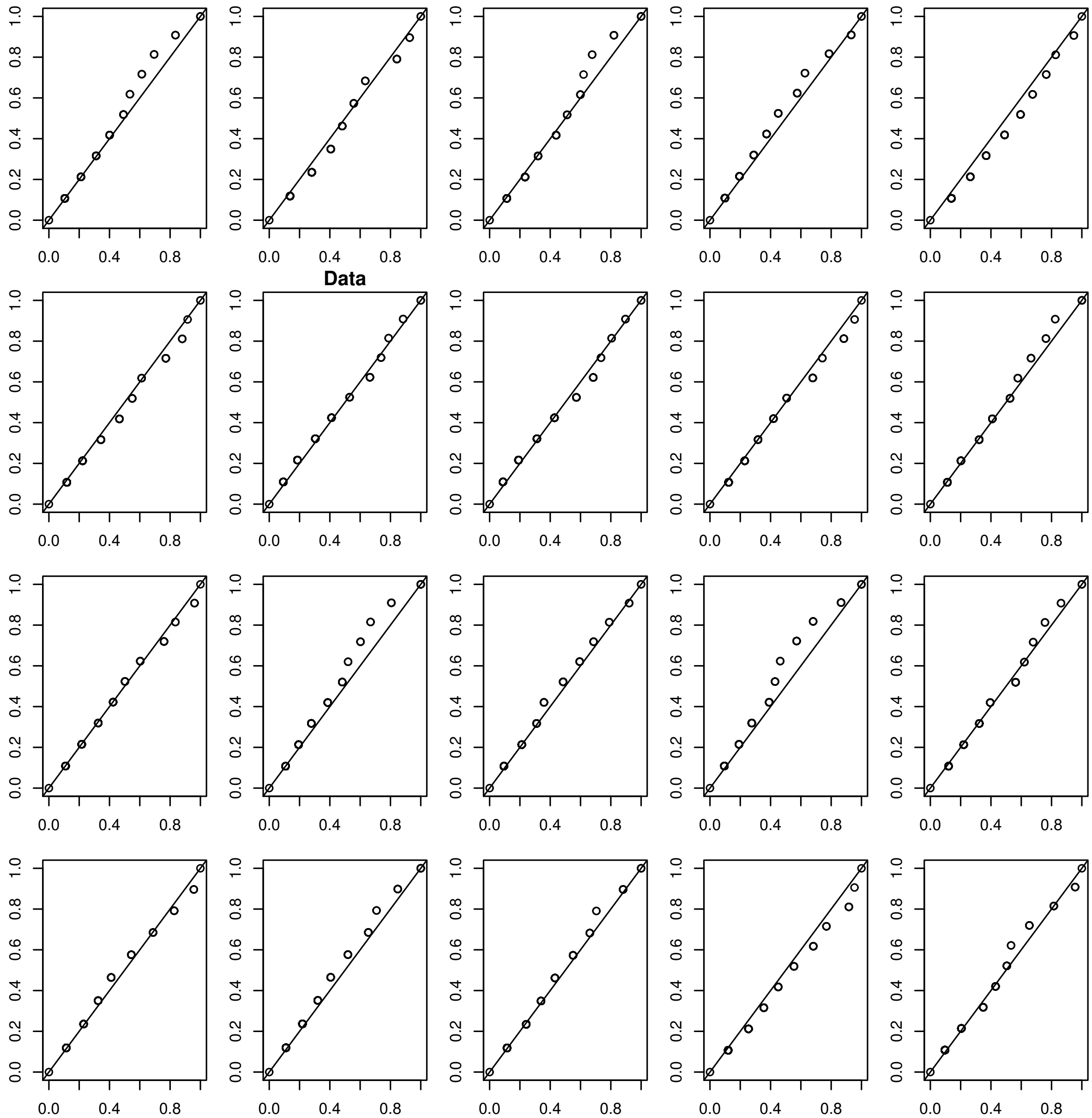

Figure 6: Line-up of Null and Observed (as indicated) TTT Plots

Figures 5 and 6 support our initial intuitions about the death rates associated with 8000m Nepalese peaks. Figure 5, which compares the raw conditional death rates (hazard probabilities), shows that the observed death rates are indistinguishable from the null

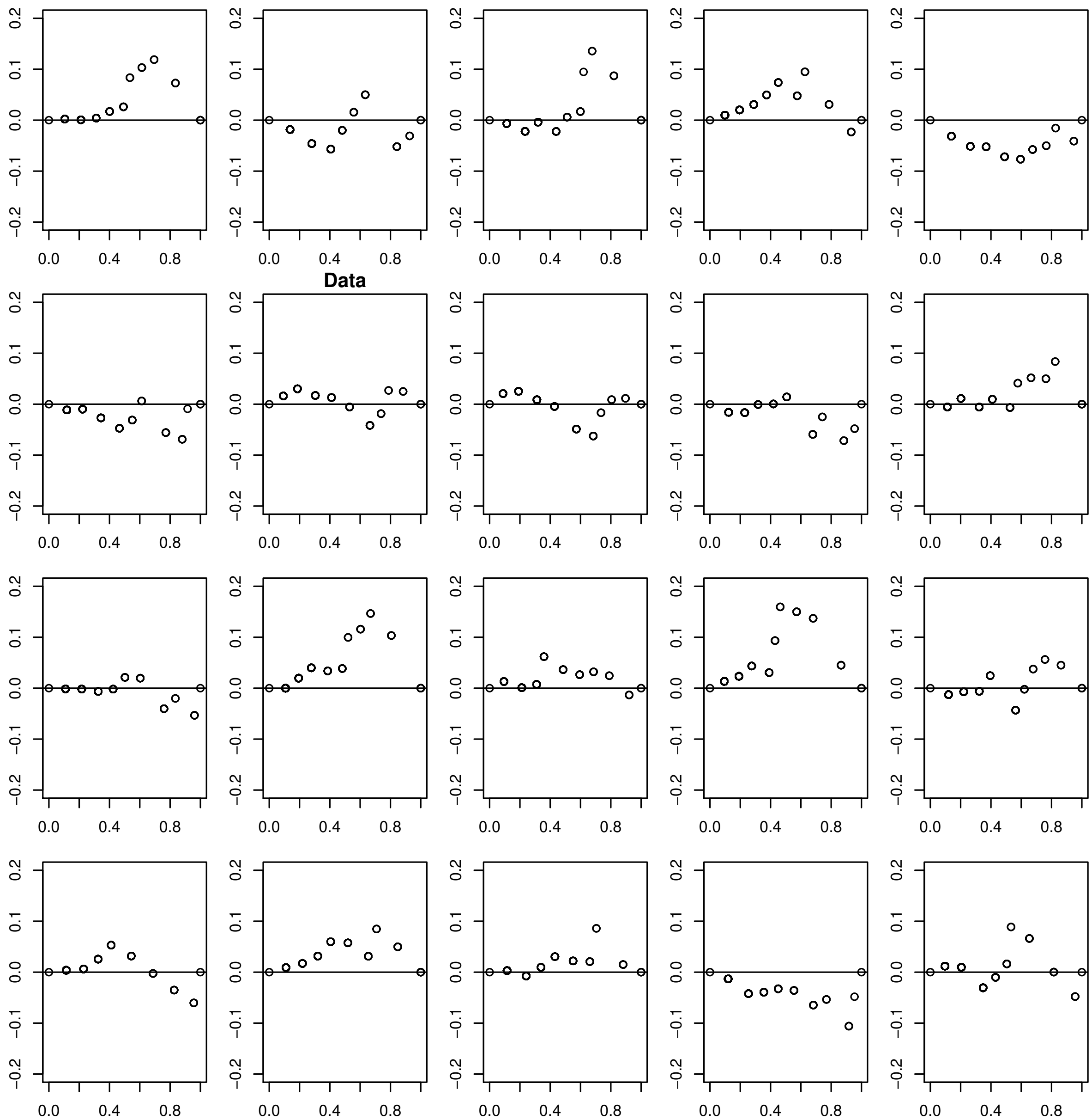

Figure 7: Line-up of TTT Plots on Vertical Scale

(constant risk) rates. These hazard plots, however, are busy and difficult to compare, so we also have the TTT plot comparison in Figure 6. Once again, the observed conditional death rates fall squarely within the null distribution. Finally, since observing deviations off a 45-degree line may be difficult, in the spirit of a Tukey mean-difference plot, Figure 7 displays rotated TTT plots with departures from the 45-degree line shown on a vertical axis. In all cases, one cannot reject the null hypothesis of a constant death risk.

5.3 TTT Plot Test Statistic

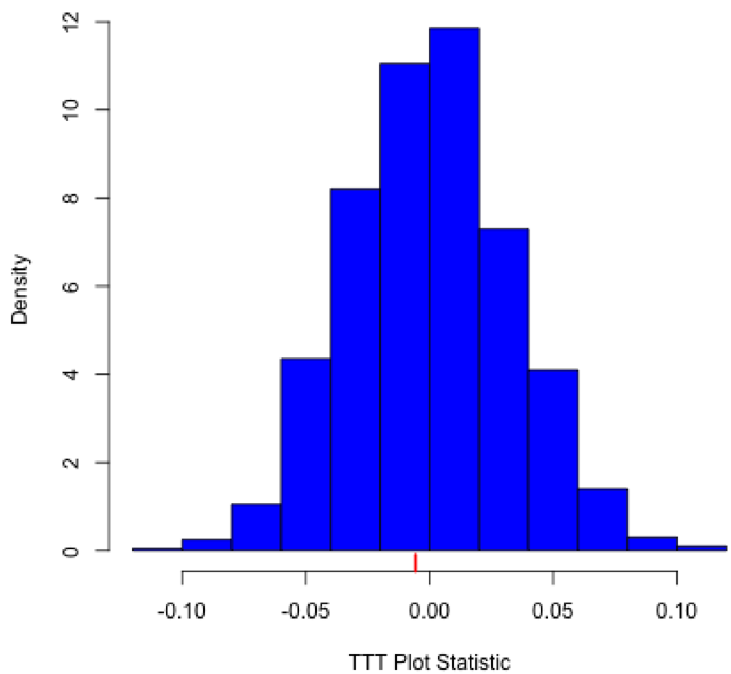

Figure 8: Observed (shown in red) and Null Distribution of TTT Plot Statistic

We can further verify the graphical inference results by using the mean negative deviation from the 45-degree line as a test statistic. To generate a null distribution for this TTT Plot Statistic, we constructed 1000 simulated datasets based on a constant death rate and calculated the statistic for each. The null distribution and the observed TTT plot statistic are illustrated in Figure 8. (Recall that decreasing death rates will produce large positive values of the TTT plot statistic.) The observed TTT plot statistic had a p-value of 0.562. We thus once again do not reject the null hypothesis of constant death rate.

5.4 Power Calculations

Naturally, the inability to reject the null hypothesis of a constant death rate does not necessarily mean that the death rate is in fact constant. And to the extent that there is no conservative position in the Roberts-Viesturs controversy, it would be unfair to side with Roberts's claim of Russian roulette based purely on the null result and without further analysis.

Figures 9 and 10 provide further support for constant death rate by giving a sense of the statistical power of the methods used above. Along with the null plots for constant death rate, these figures plot simulated data in which the death rate decreases by 10% per subsequent attempt (i.e., 1.71%, 1.53%, 1.38%...). Figure 9, which contains the traditional TTT plots, results in an equivocal hypothesis test. The 10% plot exhibits a modest convex curve, suggesting that it might be the data, but relative to the null plots, its convexity is too subtle to be a clear choice. Replotting the conventional TTT plot on a vertical scale offers some, but not definitive, improvement. In Figure 10, the 10% plot's convexity is more noticeable, although it is still uncertain whether it is "clearly" different from the others.

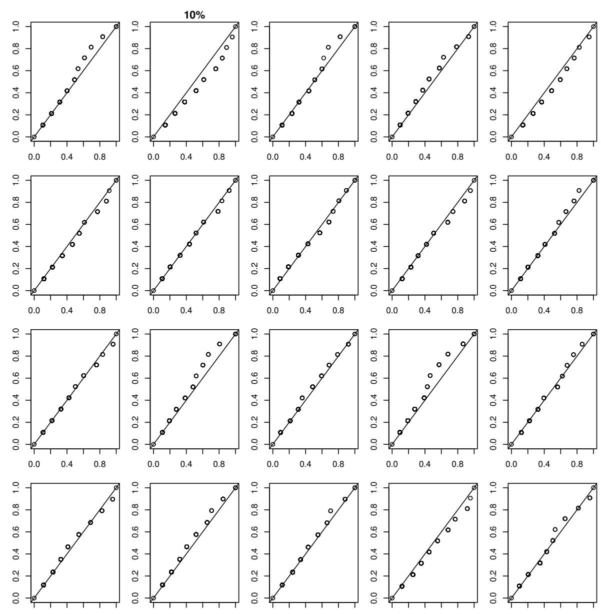

Figure 9: TTT Line-Up for 10% Decreasing Death Risk

We can arrive at a more definite answer using the TTT Plot Statistic (0.0676, p=0.02). With a p-value of 0.02, we can reject the null at the conventional 0.05 level. Thus, our

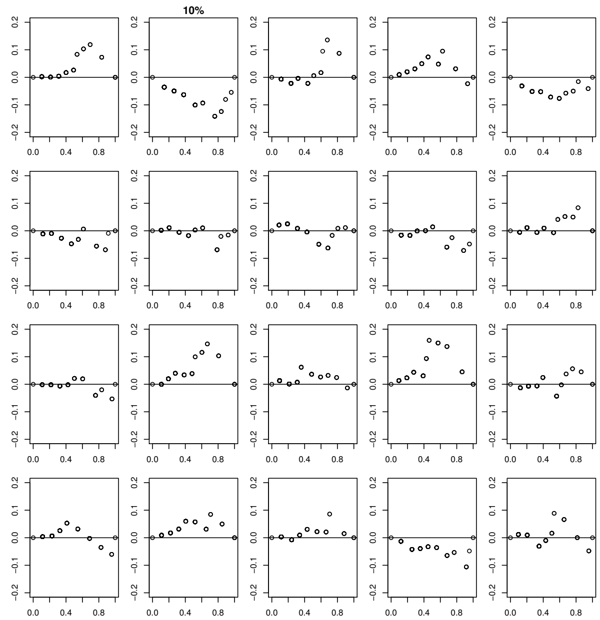

Figure 10: TTT Line-Up for 10% Decreasing Death Risk (vertical scale)

methods are capable of detecting a 10% reduction per attempt in death rate.

6 Extensions

6.1 Discrete-Time Survival Analysis

As previously mentioned, one concern with the basic death rates presented above is the potential influence of the peak climbed, the advancing state of technology, or the increasing availability of peaks like Everest to less accomplished climbers. For example, if Everest, a relatively less hazardous and technical peak, was frequently climbed first, and then Annapurna, a historically far more dangerous peak, was typically climbed later, that increase in danger might swamp any increase in safety associated with a climber's increasing experience.

Using a discrete-time survival analysis, we can control for the peak involved and the year of the attempt, minimizing the above confounding problem. One possible model is:

$$cloglog(\hat{h}(t_k)) = \sum_i \alpha_i D_i + \beta_0 Year + \sum_{j=1}^{7} \beta_j Peak_j$$

where *Year* represents the calendar year of the attempt (suitably shifted), and $Peak_j$ are a series of dummy variables for the eight different Nepalese 8000m peaks. $\sum_i \alpha_i D_i$ is the estimated baseline hazard rate, controlling for the other covariates, where once again the D_i are indicator variables for the various time periods.

Having estimated the baseline hazard rate, we can then examine its TTT plot to assess whether it is constant. Figure 11 provides the relevant TTT line-up plotted on the vertical scale, with the plot marked "Data" denoting the observed dataset.

To construct the simulated datasets, we fit a discrete-time survival analysis model with a constant baseline hazard (i.e., α_0 rather than $\sum_i \alpha_i D_i$) to the data. Then, to construct the climbing population, simulated climbers were drawn with replacement from the observed dataset, creating realistic "attempted peak" patterns. For each climb, the climber's death risk was then estimated using the constant DTSA model, and a suitable Bernoulli draw determined whether the climber died on the attempt.

As seen in Figure 11, the observed data is fully consistent with the simulated null plots. Indeed, rather than exhibiting the convexity of a decreasing death rate, it exhibits the concavity of a slight increasing trend. The p-value of the TTT Plot Statistic is 0.815, and we thus again do not reject the null hypothesis of constant death rate.

6.2 Ascent Rates

Ascent rates based on experience provide the observed null results with one final perspective. As Figure 12 readily demonstrates, experience has a dramatic effect on the probability of a successful ascent. First-time climbers on Nepalese 8000m peaks start with a 23% probability of success, which increase to 33% on their next attempt and ultimately rises to approximately 50% for veterans. Indeed, the effect is so dramatic that the increasing ascent rate is readily discerned when placed in a line-up of null plots with constant ascent rates. No TTT plots or TTT plot statistics are necessary for inference in this case.[3]

7 Discussion

So is high-altitude mountaineering like Russian roulette? While greater experience on high-altitude peaks is clearly associated with higher ascent rates, it does not appear to be associated with lower death rates. To be sure, we have principally confirmed a null result, but our methods have sufficient power to detect decreases of about 10% or greater

[3]TTT plots for ascent rates would necessarily require different calculations than those proposed in this Article, because climbers can have multiple ascents. Unlike death, success does not censor further observation of a climber.

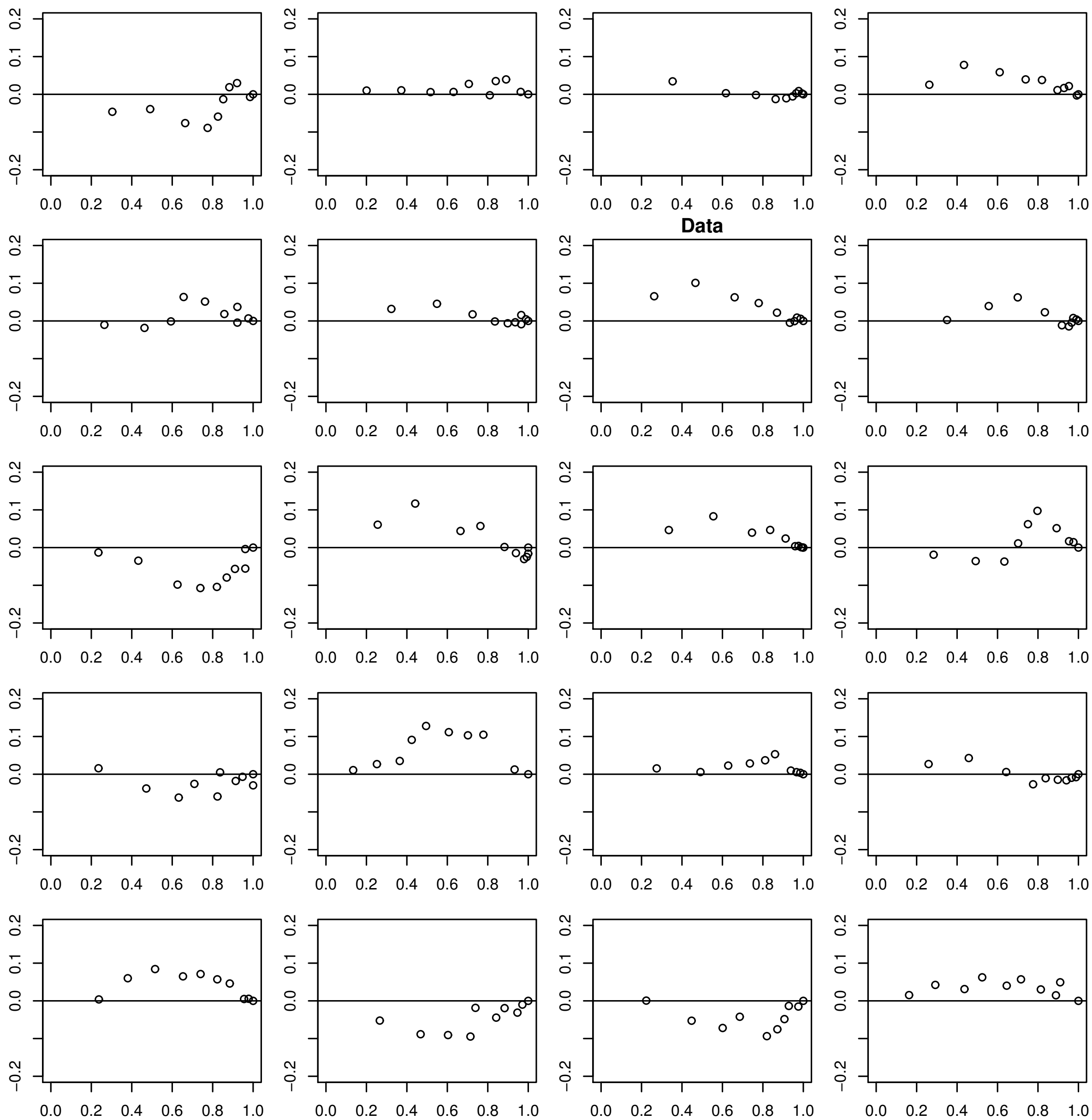

Figure 11: TTT Line-Up for Baseline Hazard Estimated Using Discrete-TIme Survival Model

per additional attempt, suggesting that if experience has any salutary effect, it is small. Regardless of experience, the objective death risk on 8000m peaks is sizable and simply

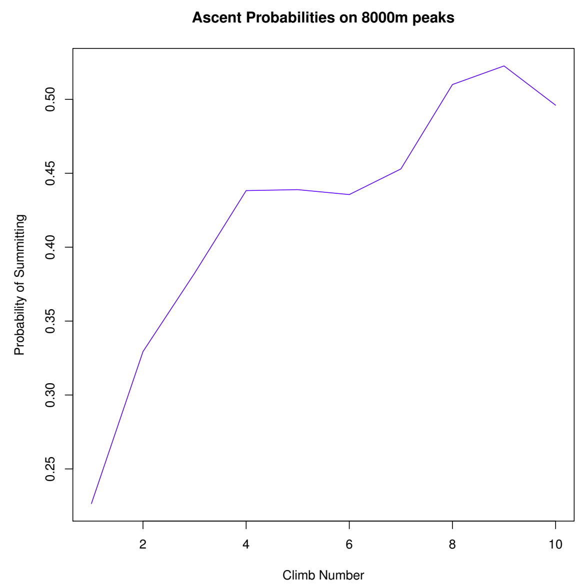

Figure 12: Ascent Rates on Nepalese 8000m Peaks

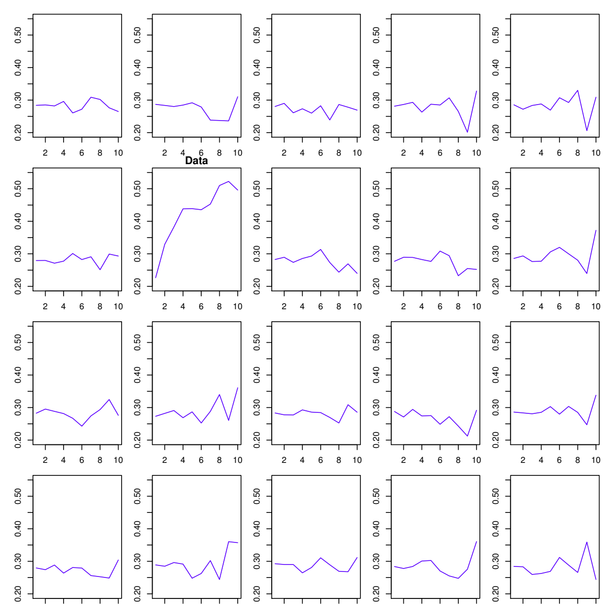

Figure 13: Line-up of Ascent Rates

does not go away. This null result remains even when controlling for the year of the attempt and the peak attempted, and is sharply contrasted with ascent rates, in which experience seems to result in dramatic improvements.

One remaining confounding issue involves the relationship between experience and risk. Conventional wisdom suggests that experience is correlated with skill and judgment, resulting in a decrease in risk, and this study has relied on that premise. However, to the extent that climbers choose whether to go on subsequent expeditions, it may be that those climbers who make repeated attempts have higher risk profiles. If so, changes or differences in risk perception and tolerance may confound whatever benefits that experience provides. Capturing this risk profile is unfortunately nearly impossible, and thus this study arguably represents the best we can do given the available data.

Finally, as with all population studies, one cannot make definitive claims about individual climbers. In response to this study, Ed Viesturs may continue to maintain that he is in a special class of mountaineers who are uniquely attentive to risk and who with each climb become ”smarter, faster, stronger, [and] more efficient.” What this study shows, however, is that at least among the general population of climbers, experience appears to have no salutary effect on the risks presented by 8000m peaks. For the typical climber, high-altitude mountaineering is Russian Roulette.

In closing, the reflective climber may ask whether the analogy to “Russian roulette” is ultimately imprecise. To the extent that David Roberts merely chose a colorful way to frame his debate with Ed Viesturs — namely, whether the risks on high-altitude peaks were controllable — this study's results suggest that Roberts is right. High-altitude mountaineering is like Russian roulette in the sense that each attempt is a static, probabilistic draw from a lottery of life and death. The vast majority of the risk is uncontrollable. Yet, “Russian roulette” often carries an additional, pejorative connotation—one of pointlessness, or risk experienced purely for its own sake. In this secondary sense, mountaineering is arguably never like Russian roulette. To purists, mountaineering is about neither the risk nor even the ascent, but rather about what difficult obstacles and extreme conditions teach climbers about themselves. The death risk may be significant and uncontrollable, but it is merely the price paid in the service of that journey.

References

Andersen, P. K., O. Borgan, R. D. Gill, and N. Keiding (1993): *Statistical Models Based on Counting Processes*, Springer.

Barlow, R. and R. Campo (1975): “Total time on test processes and applications to failure data analysis,” in R. Barlow, ed., *Reliability and Fault Tree Analysis*, SIAM, 451–481.

Bender, R., T. Augustin, and M. Blettner (2005): "Generating survival times to simulate cox proportional hazards models," *Statistics in Medicine*, 24, 1713–1723, URL http://dx.doi.org/10.1002/sim.2059.

Boyce, J. R. and D. P. Bischak (2010): "Learning by doing, knowledge spillovers, and technological and organizational change in high-altitude mountaineering," *Journal of Sports Economics*, 11, 496–532, URL http://jse.sagepub.com/content/11/5/496.abstract.

Brillinger, D. R. (2008): "The 2005 neyman lecture: Dynamic indeterminism," *Science*, 23, 48.

Buja, A., D. Cook, H. Hofmann, M. Lawrence, E.-K. Lee, D. F. Swayne, and H. Wickham (2009): "Statistical inference for exploratory data analysis and model diagnostics," *Phil. Trans. R. Soc. A*, 367, 4361–4383.

Burtscher, M., M. Philadelphy, W. Nachbauer, and R. Likar (1995): "The risk of death to trekkers and hikers in the mountains," *JAMA*, 273, 460.

Davies, P. (2008): "Approximating data," *J. Korean Stat. Soc'y*, 199–200.

Firth, P. G., H. Zheng, J. S. Windsor, A. I. Sutherland, C. H. Imray, G. W. K. Moore, J. L. Semple, R. C. Roach, and R. A. Salisbury (2008): "Mortality on mount everest, 1921-2006: descriptive study," *BMJ*, 337.

Fontanarosa, P. B., R. B. Huey, and X. Eguskitza (2000): "Supplemental oxygen and mountaineer death rates on everest and k2," *JAMA*, 284, 181.

Givens, G. H., and J. A. Hoeting (2005): *Computational Statistics*, Wiley-Interscience.

Hoyland, A. and M. Rausand (1994): *System Reliability Theory: Models and Statistical Methods*, Wiley-Interscience.

Huey, R. B. (2001): "Mountaineering in thin air," in R. C. Roach, ed., *Hypoxia: From Genes to the Bedsite*, 225.

Huey, R. B. and X. Eguskitza (2001): "Limits to human performance: elevated risks on high mountains," *Journal of Experimental Biology*, 204, 3115–3119, URL http://jeb.biologists.org/content/204/18/3115.abstract.

Huey, R. B., R. Salisbury, J.-L. Wang, and M. Mao (2007): "Effects of age and gender on success and death of mountaineers on mount everest," *Biology Letters*, 3, 498–500, URL http://rsbl.royalsocietypublishing.org/content/3/5/498.abstract.

Ilgner, A. (2003): *The Rock Warrior's Way: Mental Training for Climbers*, Desiderata.

Jolly, J. (2010): "Elizabeth hawley, unrivalled himalayan record keeper," URL http://www.bbc.co.uk/news/world-south-asia-10268549.

Pollard, A. and C. Clarke (1988): "Deaths during mountaineering at extreme altitude," *The Lancet*, 331, 1277–, URL http://linkinghub.elsevier.com/retrieve/pii/S0140673688920934.

Roberts, D. (2005): *On the Ridge Between Life and Death*, Simon & Schuster.

Salisbury, R. (2003): *The Himalayan Database: The Expedition Archives of Elizabeth Hawley*, American Alpine Club.

Salisbury, R. and E. Hawley (2007): "The himalaya by the numbers: A statistical analysis of mountaineering in the nepal himalaya," URL http://www.himalayandatabase.com/downloads/HimalayaByNbrs.pdf.

Viesturs, E. and D. Roberts (2006): *No Shortcuts to the Top: Climbing the World's 14 Highest Peaks*, Broadway.

Westhoff, J. L., T. D. Koepsell, and C. T. Littell (2012): "Effects of experience and commercialisation on survival in himalayan mountaineering: retrospective cohort study," *BMJ*, 344.

Willett, J. D. S. . J. B. (2003): *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*, Oxford.

Windsor, J. S., P. G. Firth, M. P. Grocott, G. W. Rodway, and H. E. Montgomery (2009): "Mountain mortality: a review of deaths that occur during recreational activities in the mountains," *Postgraduate Medical Journal*, 85, 316–321, URL http://pmj.bmj.com/content/85/1004/316.abstract.