Pareto Distributions in Online Casino Gambling: Sensitivity to Timeframe and Associations with Self Exclusion

Xiaolei Deng[1], Tilman Lesch[1], and Luke Clark[1,2]

[1]Centre for Gambling Research at UBC, Department of Psychology, University of British Columbia, Vancouver, B.C., Canada

[2]Djavad Mowafaghian Centre for Brain Health, University of British Columbia, Vancouver, British Columbia, Canada

This is an Author Accepted Manuscript. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article is published as: Deng X, Lesch T, Clark L. Pareto distributions in online casino gambling: sensitivity to timeframe and associations with self exclusion. *Addictive Behaviors* 2021, https://doi.org/10.1016/j.addbeh.2021.106968

Address for Correspondence

Dr Luke Clark, Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, B.C., V6T 1Z4, Canada. Tel: 001 604 827 0618. Email: luke.clark@psych.ubc.ca

Xiaolei Deng, Department of Psychology, University of British Columbia, 2136 West Mall, Vancouver, B.C., V6T 1Z4, Email: xdeng@psych.ubc.ca

Funding

This project was funded by a Research Grant from the Province of BC Ministry of Finance (Gaming Policy & Enforcement Branch) awarded to Luke Clark and Tilman Lesch, and by the core funding of the Centre for Gambling Research at UBC. LC holds a Discovery Award from the Natural Sciences and Engineering Research Council (Canada) (RGPIN-2017-04069). XD holds a UBC 4 year fellowship.

Conflict of Interest

LC is the Director of the Centre for Gambling Research at UBC, which is supported by funding from the Province of British Columbia and the British Columbia Lottery Corporation (BCLC), a Canadian Crown Corporation. The research receives separate funding from the Province of BC Ministry of Finance (Gaming Policy & Enforcement Branch) to LC and TL. The dataset for this study was provided to the research team by the BCLC under a Non-Disclosure Agreement that prohibits further data sharing. The Province of BC government and the BCLC had no further role in the design, analysis, or interpretation of the study, and impose no constraints on publishing. LC has received a speaker/travel honorarium from the National Association for Gambling Studies (Australia) and the National Center for Responsible Gaming (US), and has received fees for academic services from the National Center for Responsible Gaming (US), GambleAware (UK) and Gambling Research Exchange Ontario (Canada). He has not received any further direct or indirect payments from the gambling industry or groups substantially funded by gambling. He

has provided paid consultancy to, and received royalties from, Cambridge Cognition Ltd. relating to neurocognitive testing. TL is now employed by Deloitte Consulting GmbH. XD and TL have no further disclosures to be declared.

Abstract

The Pareto effect (also known as the 80/20 rule) describes a skewed distribution of engagement that is observed for many products. In this study, we investigated Pareto estimates for online casino gambling, and tested their association with voluntary self-exclusion (VSE) as a marker of gambling harm, and examined their sensitivity to varying time windows. We used one year of betting data from the eCasino section of a provincially-run gambling website in British Columbia, Canada. The data contained 30,920 account holders who placed at least one bet on the platform from October 2014 to September 2015. The top 20% most engaged gamblers accounted for 92% (based on total number of bets) and 90% (based on net losses) of eCasino gambling activity over the year. The top 20% of online gamblers displayed higher levels of VSE enrolment than the remaining 80% (total bets: 13% vs 7%; net losses: 16% vs 6%, respectively). Pareto estimates increased with longer time windows, from one month to one year (total bets: 80% to 92%; net losses: 81% to 90%, respectively). This accumulation was driven by the relative loyalty of the most engaged gamblers, coupled with the influx of new and more transient gamblers on a month-to-month basis. These data strengthen links between concentrations of engagement with online products and measures of harm, but also highlight the dynamic nature of these estimates. One year estimates are preferable for estimating the degree of concentration.

Keywords: internet, problem gambling, behavioural tracking, expenditure, behavioural addiction.

1. Introduction

The Pareto effect, also known as the 80/20 rule, refers to a well-known effect whereby 80% of the revenue for many goods can be attributed to the top 20% most active consumers (the "vital few") (Mizuno et al., 2008; Pinto et al., 2012; Tom et al., 2014). The present study examines this distribution of involvement for online casino gambling. In many jurisdictions, traditional forms of gambling can now be accessed online, with an ongoing debate about the relative harms associated with online gambling products (Effertz et al., 2018; Gainsbury, 2015; Papineau et al., 2018). We examine characteristics of Pareto estimates in the context of an online casino platform to understand whether the vital few are more likely to experience gambling harms, and whether these estimates are sensitive to the timeframe of the dataset.

Research on disordered gambling has traditionally emphasized how a condition that affects only a minority of the population (1 - 4%) typically accounts for a much larger segment of gambling revenue (35% to 41%, in Orford et al., 2013; Tremayne et al., 2001; Williams & Wood, 2004a). Gambling involvement, expressed in both monetary and non-monetary variables, scales dose-dependently with gambling-related problems (Brosowski et al., 2015; Currie et al., 2009; Markham et al., 2016). A recent analysis of self-reported spending on land-based gambling supported the hypothesis that people with gambling problems contribute to the concentration of gambling expenditure in a small number of highly-engaged consumers (Fiedler et al., 2019). The distributional skew (operationalized by Fiedler et al as the Gini coefficient) for slot machines was higher than the skews for other forms of gambling, and exceeded the 80/20 distribution expected from the Pareto effect. Fiedler et al (2019) argue that the degree of concentration indexes the 'addictiveness' of gambling products. An earlier analysis of the Pareto effect in a state lottery -- widely regarded as a less harmful form of gambling -- found that the most active

20% of lottery users accounted for just 65% of revenue (Clotfelter & Cook, 1990). This low estimate may reflect the non-continuous nature of lotteries imposing natural limits on consumption (Fiedler et al., 2019). Existing research on these distributional effects is hampered by a reliance on self-reported expenditure, which is known to be unreliable among gamblers (Braverman et al., 2014; Volberg et al., 2001).

Behavioural tracking of account-based data enables more accurate monitoring of gambling expenditure. In an analysis of 55 million poker transactions from over 2 million online poker players (Fiedler, 2012), the most active 10% of gamblers generated 91% of the 'rake' to the operator. Using a dataset from the European *bwin* platform, Tom, LaPlante, and Shaffer (Tom et al., 2014) calculated Pareto estimates for online sports betting and online casino games, using the distributions for the total number of bets and net loss. For fixed odds sports betting, which was the most popular mode of gambling on the platform, 80% of bets were placed by the top 18% of gamblers, and 80% of spending was attributed to just the top 7% of gamblers. Thus, in both studies (Fiedler, 2012; Tom et al., 2014), Pareto estimates for online forms of gambling exceeded the typical 80/20 distribution.

To further substantiate the link between Pareto estimates and product 'addictiveness', the study by Tom et al (2014) used a disordered gambling screening tool, the Brief Biosocial Gambling Screen (BBGS), to test whether the vital few have increased levels of disordered gambling. These data can arbitrate whether the vital few constitute a vulnerable group with gambling problems, or 'wealthy and healthy' high rollers (e.g. Zeng & Forrest, 2009). For the spending distribution, 53% of the vital few, compared to 26% of the residual group (the "trivial many"), scored in the range for disordered gambling. Unfortunately, only 2% of the website subscribers responded to the invitation to complete the BBGS, and more than a quarter of those scored in the range for disordered gambling, raising concerns regarding representativeness. The current study uses an alternative marker of gambling harm. Voluntary self-exclusion (VSE) status is an established proxy for disordered gambling in online gambling datasets, given the practical challenges with deploying screening instruments on a gambling website (Deng et al., 2019; Finkenwirth et al., 2020). VSE programs enable gamblers to bar themselves from gambling (via online or land-based facilities) for set periods of time. Approximately three-quarters of VSE individuals indicate disordered gambling on screening questionnaires (Hayer & Meyer, 2011b; McCormick et al., 2018), and VSE is further related to financial and relationship difficulties, and mental health problems (Kotter et al., 2018; Motka et al., 2018; Wardle et al., 2012).

In the present study, we examine the Pareto effect as an expression of these distributional properties, and its association with VSE status, in one year of data from the provincial online gambling platform for British Columbia (BC), Canada. We analyze data from the eCasino section that includes online slot machines and table games. The first objective was to test whether the top 20% of online gamblers displayed higher levels of engagement with the VSE program. The second, more exploratory question was to examine how sampling windows influence Pareto estimates. The Pareto estimate is a time-bound population parameter, and prior work has derived estimates using varying timeframes, e.g. six months in Fiedler (2012) versus one year in Tom et al. (2014). Time frames of assessment can exert a range of subtle influences in research on gambling (Blaszczynski et al., 2008; Brosowski et al., 2015). We reasoned that if the most highly engaged gamblers were more likely to return to a platform, Pareto estimates may be higher for datasets that cover longer time periods. We tested this by assessing the temporal stability of the Pareto estimates over an increasing window of data from one month to one year. As a reliability check, we report analyses for both total number of bets and net loss, as non-monetary and monetary expressions (respectively) of gambling intensity (see also Tom et al., 2014).

2. Methods

2.1. Participants

The sample consisted of 30,902 account holders on the BC PlayNow.com online platform, who placed at least one bet in the 'eCasino' section of the website between 1 October 2014 and 30 September 2015. PlayNow.com is operated by the British Columbia Lottery Corporation (BCLC) (since 2004) and is the only regulated online gambling website in the province; access is restricted to BC residents. The eCasino section comprises online slot machine games, table games (e.g. roulette, blackjack), video poker, and other probability games. We have previously reported that the eCasino section accounted for 97% of bets placed on the PlayNow platform (Lesch & Clark, 2020). Within the eCasino, slot machines accounted for 73.8% of bets and 75.6% of spend (see Supplementary Table 2). In our sample, 2,465 users (7.98%) had a record of enrolment in the BC voluntary self-exclusion program since opening their account. However, based on the records we received, we could not differentiate users who had self-excluded prior to the year under study (and whose self-exclusion had expired) from users who had enrolled in self-exclusion during the year. We could also not differentiate people by length of enrollment. In British Columbia, people can self-exclude for six months, one year, two years, or three years.

2.2. Procedures

The study protocol was approved by the Behavioural Research Ethics Board at the University of British Columbia. As secondary data, we did not obtain direct consent; rather the user agreement upon registering with PlayNow.com stated that customers' gambling data may be shared with third parties for research purposes. All betting-related variables, such as date, time, product, bet amount, win amount, etc., were automatically recorded by PlayNow.com. Prior to data transfer, the BCLC Data Analytics team assigned each user a unique and randomly generated ID number that was linked to their betting behaviour. The data was stored on a secure server (Westgrid) hosted in Canada and all data aggregations and analyses were performed on the secure server through Datagrip, a Structured Query Language empowered database software. Data were transformed into fixed point format prior to data analyses to avoid rounding errors.

The Pareto estimates were calculated for two alternative distributions: i) the distribution of the total number of bets per user, ii) the distribution of the cumulative net loss per user. For the net loss distribution, the net loss for each bet was first calculated by subtracting any amount won from the bet size, such that a positive number indicates net loss. These net losses were then aggregated for each user over the time period of the analysis. For both Pareto calculations, we rank-ordered all users in the distribution, and we calculated the sum totals for the entire distribution, and for the top 20% of the distribution (6,180 of 30,902 individuals). Thus, the Pareto estimate for total number of bets was the percentage of the grand total of bets generated by the top 20% of users. Calculation of the Pareto estimate for net loss was complicated by the fact that some users recorded a net gain within the time period (n = 3,139 (10.16%) for the full year, compared to n = 1,095 (3.54%) who broke even, and n = 26,668 (86.3%) who recorded a net loss, see Supplementary Table 1). With a sufficient amount of net gain, the top 20% can account for more than 100% of activity[1]. It is counterintuitive for a Pareto estimate to exceed

[1] As an example, assume the total loss of all users is 120 dollars, the loss of the top 20% users is 100 dollars, and the total winnings of all users is 30 dollars. In this scenario, the net loss of all

100%, so to impose an upper bound of 100% on the Pareto estimate, our primary analysis excluded these net gain values from the grand total, effectively re-coding these values as zeros. Using the full sample for both distributions also facilitates comparison of the Pareto estimates and VSE proportions for the two variables. To test the impact of this decision, the Supplementary Material shows the re-calculated Pareto estimates for net loss when i) we include these net gain amounts, and ii) when we reduce the overall distribution to only include the net losers (i.e. the top 20% becomes 5,334 of 26,668 users).

We used chi-squared tests to determine whether the top 20% were more likely to have a VSE flag than the remaining 80%. The 95% confidence intervals were constructed around the difference in the VSE proportions by the top 20% and the remaining 80%.

3. Results

Our sample of 30,902 users generated over half a billion individual bets and sustained a total net loss of over 55 million dollars over the year. Month by month activity and yearly totals are shown in Table 1, and further one year descriptives are available in Supplemental Table 1. It is unclear if the minor peak in activity in July 2015 reflects a regular seasonal cycle or a response to external events, such as marketing.

---

users is 90 dollars, which is smaller than the loss incurred by the top 20%, leading to a Pareto estimate of 111%. The risk of this inflation is also greater for short timeframes, see Supplemental Table 4.

Table 1. Monthly activity statistics for the one year dataset

Month	Active Users	Numbers of Bets	Net Loss
October 2014	11,456	42,835,566	4,403,086
November 2014	12,710	43,578,319	4,030,561
December 2014	11,808	45,603,331	4,792,050
January 2015	12,551	46,421,285	4,143,436
February 2015	12,481	43,814,648	4,176,757
March 2015	13,134	46,931,989	4,037,562
April 2015	12,514	47,771,336	4,651,408
May 2015	12,690	49,119,381	4,623,319
June 2015	12,384	46,023,852	4,529,954
July 2015	14,305	53,276,023	5,212,882
August 2015	12,896	50,732,714	5,383,193
September 2015	13,694	48,454,865	5,201,167
Grand Total	30,902	564,563,309	55,185,375

3.1. Rates of VSE in the Pareto Subgroups

Among the top 20% of users ranked on total numbers of bets (n = 6,180), 832 users (13.46%) had a VSE flag. In the remaining 80% (n = 24,722), there were 1,633 users (6.61%) with a VSE flag. The proportion with a VSE record was significantly higher in the top 20% compared to the remaining 80% (difference = 6.85%, 95% CI = 5.94% - 7.76%, $\chi^2(1) = 316.71$, $p < 0.001$).

For the distribution of net losses, among the top 20% of users (n = 6,180), 1,007 users (16.29%) had a VSE flag. In the remaining 80% of users (n = 24,722), there were 1,458 users (5.90%) with a VSE flag. Again, this proportion was significantly higher in the vital few (difference = 10.39%, 95% CI = 9.42% - 11.36%, $\chi^2(1) = 755.54$, $p < .001$).

3.2. Stability of the Pareto Estimates by Timeframe

For total number of bets, the Pareto estimate based on active users within any single calendar month varied from 78.45% to 81.95% (see Figure 1a and Supplemental Table 5). By contrast, the Pareto estimate for the full year dataset was 91.8%. As shown in Figure 1a and Supplemental Table 3, accumulative Pareto estimates increased steadily with the length of the data window, from 80.29% on the first month of data (i.e. October 2014) to the 91.75% value for the entire year. Visual inspection of the figure indicates some evidence of stabilizing at around 12 months.

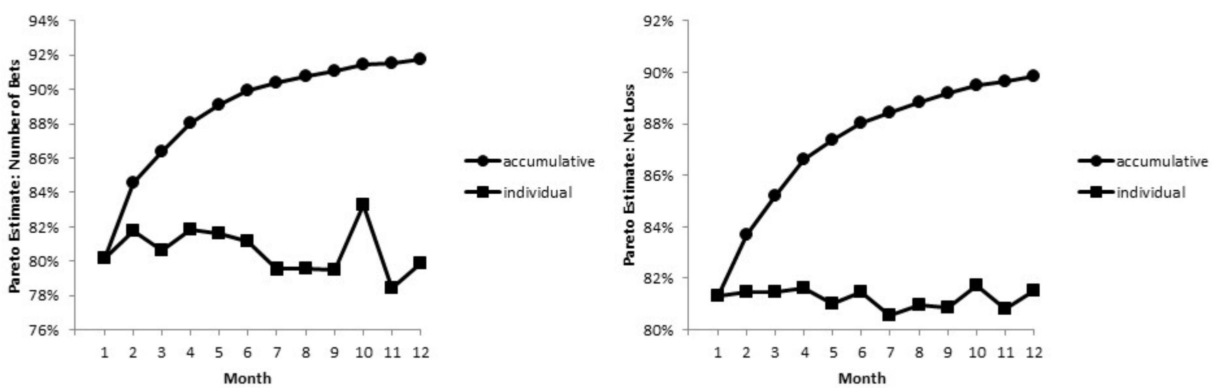

Figure 1. Pareto estimates (i.e. activity attributable to the top 20% most engaged gamblers, plotted month by month ('individual') or as a function of an expanding time window ('accumulative'). Left: from the distribution of the total number of bets; right: from the distribution of the total net loss.

A similar pattern was observed for Pareto estimates based on the net loss. The estimates by single calendar month varied from 80.6% to 81.7% (see Figure 1b and Supplemental Table 5), whereas the Pareto estimate for the full year dataset was 89.9%. The accumulative Pareto estimates increased steadily from 81.3% in the first month to the full year estimate (see Figure 1b and Supplemental Table 4).

The one year Pareto estimates were comparable when the dataset was restricted to the online slot machine products (see Supplemental Table 2). The accumulation effect by timeframe was also observed in the top 1% most active users (see Supplemental Table 3 and Table 4).

Why does the Pareto estimate vary so much with length of data window? We tested two underlying mechanisms, noting that these effects are not mutually exclusive. One possibility is that the users in the top 20% are more likely to return the platform in the subsequent months. We tested for this differential 'loyalty' of users, by calculating the percentage of users from Month 1 who placed bets in the eCasino in each subsequent month. The top 20% of users were indeed more loyal to the platform, with approximately 80% still returning after 12 months, compared to approximately 60% in the remaining users (see Figure 2a). We note that these loyal customers do not necessarily continue to rank in the top 20% in each subsequent month. This difference was confirmed by permutation test ($p < 0.01$).

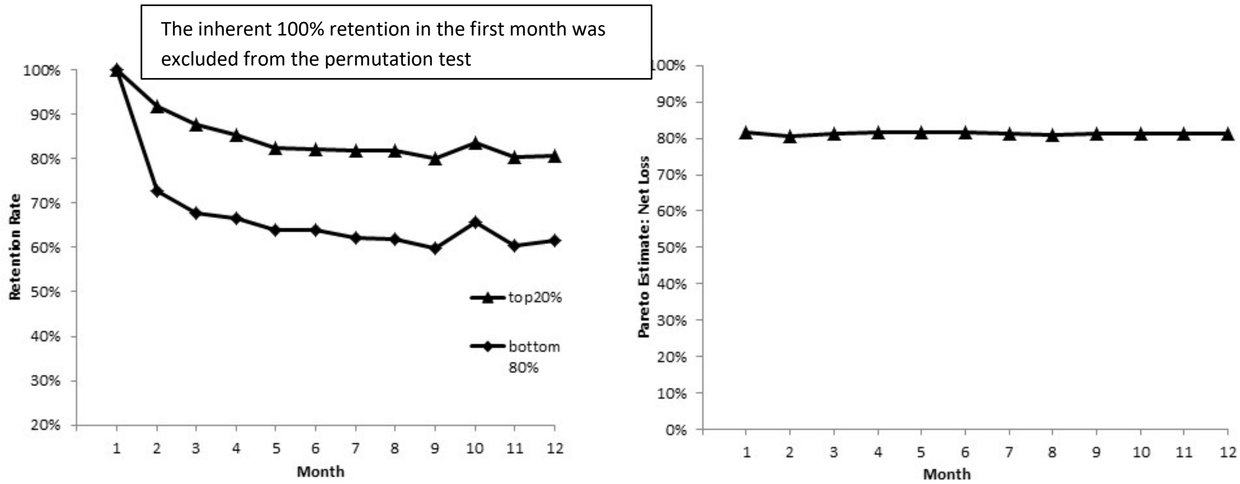

Figure 2. Left: the proportion of gamblers in the top 20% and residual 80%, based on the net loss distribution, who return to the platform in each subsequent month. This figure illustrates the differential loyalty of the 'vital few' over subsequent months. Right: the effect of the expanding time window on the Pareto estimates for net loss, when restricted to the gamblers who were active in the first month.

A flipside to the greater loyalty in the top 20% is an influx of new and potentially more transient customers, on a month-to-month basis. To test the impact of this influx, we excluded new customers from the accumulation analysis for net loss: we calculated the accumulating Pareto estimates following *only* the users who gambled in the first month (n = 11,456). In this analysis (see Figure 2b), the accumulating Pareto estimates remain remarkably stable, between 80.65% and 81.51%. By inference, the accumulating effect in the main analysis is at least partly attributable to the month-to-month influx of users that was excluded here.

4. Discussion

Account-based data from online gambling offer a powerful means of tracking behaviour from individual gamblers over time. In this study, we explored distributions of involvement in a large dataset from the eCasino section of a Canadian provincial online gambling platform. In the one year dataset, we observed that the top 20% in the distributions for total bets and net loss accounted for 92% and 90% of activity, respectively. These estimates exceed the conventional Pareto (80/20) rule. The top 20% of online casino gamblers displayed higher levels of VSE enrolment than the remaining 80%. We observed a striking increase in the Pareto estimates as a function of timeframe, from about 80% in one month of activity to about 90% in the full year. This instability appeared to be driven by the relative loyalty of the top 20% of gamblers over time, coupled with the influx of new and more transient gamblers on a month-to-month basis.

In our study, VSE status served as a marker for disordered gambling. Overall, 8% of our sample had a record of enrollment in the provincial VSE program. Our study demonstrated that the VSE rate~~s~~ among the top 20% of gamblers was approximately doubled (compared to the remaining 80%) for the distribution of total bets (13% vs 7%), and almost tripled for the distribution of net loss (16% vs 6%). There is an indication here that the monetary variable (net loss) may be more sensitive to VSE than the non-monetary variables (total bets), although our analysis plan did not test for this formal interaction, and we note that prior research is inconsistent regarding these differential sensitivities (Dragicevic et al., 2015; Finkenwirth et al., 2020; Gray et al., 2012). These findings corroborate the conclusion by Tom et al (Tom et al., 2014) using the BBGS screener, that the vital few on an online gambling platform comprise an elevated numbers of problem gamblers, and challenge any argument that the vital few are merely ‘healthy and wealthy’ high rollers (see also Zeng & Forrest, 2009). Our dataset used VSE records for all active gamblers in the period under investigation, thus addressing the sampling bias in Tom et al in which only 2% of bwin subscribers submitted the BBGS.

This increased rate of VSE supports the assertion that the "vital few" do manifest indicators of disordered gambling. VSE programs have been implemented across many international jurisdictions (Motka et al., 2018), and enrolling in VSE offers a concrete step towards controlling excessive gambling. However, VSE also has limitations as a marker of disordered gambling (see also Finkenwirth et al., 2020). VSE status in the current study was a lifetime record, and we cannot distinguish gamblers who self-excluded during the period under investigation from those with historic records prior to October 2014. A prospective study of future VSE can help resolve this limitation and shed light on online behaviours that predate VSE enrolment. More broadly, there are recognized caveats with VSE as a 'proxy' for disordered gambling, given that not all VSE gamblers appear to display gambling problems, and many problem gamblers evidently do not utilize VSE (Finkenwirth et al., 2020). Although VSE was significantly elevated in the top 20% in our data, the rates were still low in both subgroups. Nevertheless, our results support an argument that Pareto distributions provide an indicator of the harmful potential of gambling products.

Our observed Pareto estimates in the full year online gambling dataset exceed the classic Pareto 80/20 rule, and are broadly consistent with other Pareto estimates for forms of online gambling (Fiedler, 2012; Tom et al., 2014). These estimates are markedly higher than revenue attributable to the vital few in an analysis of land-based lottery sales (Clotfelter & Cook, 1990), widely viewed as a less harmful form of gambling (Costes et al., 2018). These data support the argument (Fiedler et al., 2019) that greater concentrations of consumption (i.e. *in excess* of the standard 80/20 rule) may be the sign of an addictive product. Whereas Fiedler et al tested this hypothesis using survey data based on self-reported gambling expenditure, our analyses use tracked data of online gambling activity. Nevertheless, further work is required to explore this hypothesis. We note that distributional analyses for consumption of drugs of abuse can approximate the 80/20 rule (the top 20% generate around 72% of alcohol sales (Stockwell et al., 2009) and 82% of cannabis consumption(Callaghan et al., 2019)). Fiedler et al speculate that physiological limits on consumption of alcohol and other drugs may inherently constrain the degree of distributional skew. By extension, their hypothesis may apply primarily to gambling and other technological addictions, given the potential for continuous consumption. We confirmed our Pareto estimates for the online slot machine products specifically, as the modal form in the eCasino, but future research comparing different online gambling product types is warranted. We also recognize that other modelling techniques exist for characterizing distributional properties, including the Gini coefficient (Fiedler et al., 2019; Pinto et al., 2012).

The accumulating nature of our Pareto estimates by timeframe was observed on the distributions for total bets and net loss. A comparable effect was seen in the top 1% of users; for example, the net loss value rises from 24% in one month to 30% in one year (Supplemental Table 4). The effect could be attributed to two factors. The top 20% in Month 1 showed greater retention on the platform over subsequent months. This tendency to return to the platform could reflect between-session chasing as a key feature of disordered gambling. A recent 5 year longitudinal analysis showed that chasing was the diagnostic item that best predicted increases in gambling severity over a 5 year period (Sleczka & Romild, 2021). A consequence of the differential 'loyalty' is that the vital few account for a greater proportion of activity in larger slices of data. Future analyses may test to what extent this result depends on the breadth of gambling involvement, another metric that is associated with disordered gambling (LaPlante et al., 2014; Papineau et al., 2018; Philander & MacKay, 2014), but is likely to be under-estimated in our data due to the focus on only the eCasino section. The second factor in our data was the influx of new gamblers in later months. When these users were excluded from our accumulation analyses, the increase in the Pareto estimates with timeframe was blocked (see Figure 2b). Our interpretation is that the regular influx of transient gamblers on a month-to-month basis also contributes to the accumulation effect in the main analysis. While these dual influences of loyalty and influx may be considered 'two sides of the same coin', we tested them in separate analyses, and we cannot be sure to what extent these two influences *fully* account for the increase in the Pareto from about 80% to about 90%, or whether further factors may also play a role.

We do not regard this observed temporal instability of Pareto estimates as a flaw for these metrics: Pareto estimates provide a snapshot of the distribution of involvement with a product, but knowledge users should be aware of their sensitivity to statistical and temporal properties of the data. Our results highlight a need to use consistent timeframes in order to compare Pareto estimates across different forms of gambling or different jurisdictions (i.e. to compare 'apples with apples'). In the current study, the cumulative Pareto estimates showed some evidence of stabilizing at around 90%, indicating that a one year timeframe may be optimal for future work (see also Fiedler et al., 2019). Future research deriving Pareto estimates over longer timeframes may be undertaken to confirm our findings.

Some further limitations should be noted. First, although account-based tracking enables researchers to overcome inaccuracies in gamblers' self-reported expenditure, it is possible that gamblers may hold accounts on other online platforms (e.g. unregulated websites, in the jurisdiction under study) and also engage in land-based gambling to varying degrees. Accounting for these other forms could further influence Pareto estimates. Second, our analyses did not include or control for age and gender as important demographic factors that could be associated with either eCasino preferences or VSE (Dragicevic et al., 2015; Hayer & Meyer, 2011a). Third, net loss is an objective measure but remains subject to the game's volatility, and other metrics adjust the amount wagered by the house edge (e.g. 'theoretical loss', Auer & Griffiths, 2013). Further progress in this area would be facilitated through increased access to other online gambling datasets in the research community.

In conclusion, our findings are congruent with a contemporary public health perspective on gambling, and provide insights into the distributions of engagement and harm in the specific context of online casino gambling. In a one year dataset, we attribute roughly 90% of activity to the top 20% of users, and on a month-to-month basis, our analyses begin to characterize the patterns of movement in and out of the vital few. Given the association with VSE, one policy implication is that online gamblers falling consistently in the top part of the distribution could be excluded from marketing activities. Distributional analyses may have particular utility in the context of online gambling and other forms of online spending because they can be calculated with relative ease and objectivity using account-linked data. Our data substantiate a recent hypothesis that concentrations of activity in excess of the classical 80/20 rule observed for many goods (convenience stores, book sales, etc; Mizuno et al., 2008; Pinto et al., 2012) may be useful in assessing the harm profile of online gambling products.

References

Auer, M., & Griffiths, M. D. (2013). An empirical investigation of theoretical loss and gambling intensity. *Journal of Gambling Studies*, *30*(4), 879–887. https://doi.org/10.1007/s10899-013-9376-7

Blaszczynski, A., Ladouceur, R., Goulet, A., & Savard, C. (2008). Differences in monthly versus daily evaluations of money spent on gambling and calculation strategies. *Journal of Gambling Issues*, *21*, 98. https://doi.org/10.4309/jgi.2008.21.9

Braverman, J., Tom, M. A., & Shaffer, H. J. (2014). Accuracy of self-reported versus actual online gambling wins and losses. *Psychological Assessment*, *26*(3), 865–877. https://doi.org/10.1037/a0036428

Brosowski, T., Hayer, T., Meyer, G., Rumpf, H. J., John, U., Bischof, A., & Meyer, C. (2015). Thresholds of probable problematic gambling involvement for the German population: Results of the Pathological Gambling and Epidemiology (PAGE) Study. *Psychology of Addictive Behaviors*, *29*(3), 794–804. https://doi.org/10.1037/adb0000088

Callaghan, R. C., Sanches, M., Benny, C., Stockwell, T., Sherk, A., & Kish, S. (2019). Who consumes most of the cannabis in Canada? Profiles of cannabis consumption by quantity. *Drug and Alcohol Dependence*, *205*, 107587. https://doi.org/10.1016/j.drugalcdep.2019.107587

Clotfelter, C. T., & Cook, P. J. (1990). On the Economics of State Lotteries. *Journal of Economic Perspectives*, *4*(4), 105–119. https://doi.org/10.1257/jep.4.4.105

Costes, J. M., Kairouz, S., Monson, E., & Eroukmanoff, V. (2018). Where lies the harm in lottery gambling? A portrait of gambling practices and associated problems. *Journal of Gambling Studies*, *34*(4), 1293–1311. https://doi.org/10.1007/s10899-018-9761-3

Currie, S. R., Miller, N., Hodgins, D. C., & Wang, J. (2009). Defining a threshold of harm from gambling for population health surveillance research. *International Gambling Studies*, *9*(1), 19–38. https://doi.org/10.1080/14459790802652209

Deng, X., Lesch, T., & Clark, L. (2019). Applying Data Science to Behavioral Analysis of Online Gambling. *Current Addiction Reports*, *6*(3), 159–164. https://doi.org/10.1007/s40429-019-00269-9

Dragicevic, S., Percy, C., Kudic, A., & Parke, J. (2015). A descriptive analysis of demographic and behavioral data from Internet gamblers and those who self-exclude from online gambling platforms. *Journal of Gambling Studies*, *31*(1), 105–132. https://doi.org/10.1007/s10899-013-9418-1

Effertz, T., Bischof, A., Rumpf, H. J., Meyer, C., & John, U. (2018). The effect of online gambling on gambling problems and resulting economic health costs in Germany. *European Journal of Health Economics*, *19*(7), 967–978. https://doi.org/10.1007/s10198-017-0945-z

Fiedler, I. (2012). The Gambling Habits of Online Poker Players. *Journal of Gambling Business and Economics*, *6*(1), 1–30. https://doi.org/10.2139/ssrn.1908161

Fiedler, I., Kairouz, S., Costes, J. M., & Weißmüller, K. S. (2019). Gambling spending and its concentration on problem gamblers. *Journal of Business Research*, *98*, 82–91. https://doi.org/10.1016/j.jbusres.2019.01.040

Finkenwirth, S., MacDonald, K., Deng, X., Lesch, T., & Clark, L. (2020). Using machine learning to predict self-exclusion status in online gamblers on the PlayNow.com platform in British Columbia. *International Gambling Studies*, 1–45. https://doi.org/10.1080/14459795.2020.1832132

Gainsbury, S. M. (2015). Online Gambling Addiction: the Relationship Between Internet Gambling and Disordered Gambling. *Current Addiction Reports*, *2*(2), 185–193. https://doi.org/10.1007/s40429-015-0057-8

Gray, H. M., Laplante, D. A., & Shaffer, H. J. (2012). Behavioral characteristics of Internet gamblers who trigger corporate responsible gambling interventions. *Psychology of Addictive Behaviors*, *26*(3), 527–535. https://doi.org/10.1037/a0028545

Hayer, T., & Meyer, G. (2011a). Internet self-exclusion: characteristics of self-excluded gamblers and preliminary evidence for its effectiveness. *International Journal of Mental Health and Addiction*, *9*(3), 296–307. https://doi.org/10.1007/s11469-010-9288-z

Hayer, T., & Meyer, G. (2011b). Self-Exclusion as a Harm Minimization Strategy: Evidence for the Casino Sector from Selected European Countries. *Journal of Gambling Studies*, *27*(4), 685–700. https://doi.org/10.1007/s10899-010-9227-8

Kotter, R., Kräplin, A., & Bühringer, G. (2018). Casino Self- and Forced Excluders' Gambling Behavior Before and After Exclusion. *Journal of Gambling Studies*, *34*(2), 597–615. https://doi.org/10.1007/s10899-017-9732-0

LaPlante, D. A., Nelson, S. E., & Gray, H. M. (2014). Breadth and depth involvement: Understanding Internet gambling involvement and its relationship to gambling problems.

Psychology of Addictive Behaviors, *28*(2), 396–403. https://doi.org/10.1037/a0033810

Lesch, T., & Clark, L. (2020). *Online Gambling in British Columbia: a Descriptive Report on Gambling Participation on the PlayNow.com Platform*. https://doi.org/10.5683/SP2/LU84ZE

Markham, F., Young, M., & Doran, B. (2016). The relationship between player losses and gambling-related harm: Evidence from nationally representative cross-sectional surveys in four countries. *Addiction*, *111*(2), 320–330. https://doi.org/10.1111/add.13178

McCormick, A. V., Cohen, I. M., & Davies, G. (2018). Differential effects of formal and informal gambling on symptoms of problem gambling during voluntary self-exclusion. *Journal of Gambling Studies*, *34*(3), 1013–1031. https://doi.org/10.1007/s10899-018-9743-5

Mizuno, T., Toriyama, M., Terano, T., & Takayasu, M. (2008). Pareto law of the expenditure of a person in convenience stores. *Physica A: Statistical Mechanics and Its Applications*, *387*(15), 3931–3935. https://doi.org/10.1016/j.physa.2008.01.059

Motka, F., Grüne, B., Sleczka, P., Braun, B., Örnberg, J. C., & Kraus, L. (2018). Who uses self-exclusion to regulate problem gambling? A systematic literature review. *Journal of Behavioral Addictions*, *7*(4), 903–916. https://doi.org/10.1556/2006.7.2018.96

Orford, J., Wardle, H., & Griffiths, M. (2013). What proportion of gambling is problem gambling? Estimates from the 2010 British Gambling Prevalence Survey. *International Gambling Studies*, *13*(1), 4–18. https://doi.org/10.1080/14459795.2012.689001

Papineau, E., Lacroix, G., Sévigny, S., Biron, J. F., Corneau-Tremblay, N., & Lemétayer, F. (2018). Assessing the differential impacts of online, mixed, and offline gambling. *International Gambling Studies*, *18*(1), 69–91. https://doi.org/10.1080/14459795.2017.1378362

Philander, K. S., & MacKay, T. L. (2014). Online gambling participation and problem gambling severity: Is there a causal relationship? *International Gambling Studies*, *14*(2), 214–227. https://doi.org/10.1080/14459795.2014.893585

Pinto, C. M. A., Mendes Lopes, A., & Machado, J. A. T. (2012). A review of power laws in real life phenomena. *Communications in Nonlinear Science and Numerical Simulation*, *17*(9), 3558–3578. https://doi.org/10.1016/j.cnsns.2012.01.013

Sleczka, P., & Romild, U. (2021). On the stability and the progression of gambling problems: longitudinal relations between different problems related to gambling. *Addiction*, *116*(1), 116–125. https://doi.org/10.1111/add.15093

Stockwell, T., Zhao, J., & Thomas, G. (2009). Should alcohol policies aim to reduce total alcohol consumption? New analyses of Canadian drinking patterns. *Addiction Research and Theory*, *17*(2), 135–151. https://doi.org/10.1080/16066350801902467

Tom, M. A., LaPlante, D. A., & Shaffer, H. J. (2014). Does Pareto Rule Internet Gambling? Problems Among The "Vital Few" & "Trivial Many." *Journal of Gambling Business and Economics*, *8*(1), 73–100. https://ideas.repec.org/a/buc/jgbeco/v8y2014i1p73-100.html

Tremayne, K., Masterman-Smith, H., & McMillen, J. (2001). *Survey of The Nature and Extent of Gambling and Problem Gambling in the ACT*.

Volberg, R. A., Gerstein, D. R., Christiansen, E. M., & Baldridge, J. (2001). Assessing self-reported expenditures on gambling. *Managerial and Decision Economics*, *22*(1–3), 77–96. https://doi.org/10.1002/mde.999

Wardle, H., Griffiths, M. D., Orford, J., Moody, A., & Volberg, R. (2012). Gambling in Britain: A Time of Change? Health Implications from the British Gambling Prevalence Survey 2010. *International Journal of Mental Health and Addiction*, *10*(2), 273–277. https://doi.org/10.1007/s11469-011-9319-4

Williams, R. J., & Wood, R. T. (2004). The proportion of gaming revenue derived from problem gamblers: examining the issues in a Canadian context. *Analyses of Social Issues and Public Policy*, *4*(1), 33–45. https://doi.org/10.1111/j.1530-2415.2004.00033.x

Zeng, Z., & Forrest, D. (2009). High rollers from mainland China: A profile based on 99 cases. *UNLV Gaming Research & Review Journal*, *13*(1), 29–43. http://ovidsp.ovid.com/ovidweb.cgi?T=JS%7B&%7DPAGE=reference%7B&%7DD=psyc6%7B&%7DNEWS=N%7B&%7DAN=2009-05464-002

Supplementary Materials

Supplemental Table 1. Sample Descriptives

	Total	Net loss (top 20%)	Net loss (other 80%)	Net win	Break even
Number of users	30,902	5,334	21,334	3,139	1,095
Total net loss (CAD)	55,185,374.07	53,121,310.46	5,970,849.36	-3,906,785.75	0
Total number of bets	564,563,309	429,649,652	101,277,556	33,561,338	74,763
Mean number of bets per user	18,269	80,549	4,747	10,692	68
SD bets per user	217015.2	100,698.2	13,801.14	153319.1	303.72
Median number of bets per user	760	47,152	432	339	20
Mean wager per bet (CAD)	4.33	7.74	3.25	6.63	2.25
SD wager per bet	13.89	14.81	7.71	19.40	3.25
Median wager per bet (CAD)	1.42	1.98	1.26	2.00	1.00

Mean net loss per user (CAD)	1,785.82	9,959.00	279.87	-1,244.60	0
SD net loss per user	8749.39	15,473.34	408.56	14,373.55	0
Median net loss per user (CAD)	99.44	4,820.30	86.86	-37	0

Supplemental Table 2. One year descriptives and Pareto estimates restricted to the slot machine products on the platform, as the modal product category

Variable	Slot machines
Total users	23,240
Number of bets	416,477,565 (73.77%)
Net loss ($)	41,692,510.77 (75.55%)
Average net loss per bet ($)	0.10
Pareto estimate (total number of bets)	89.07%
Pareto estimate (net loss)	91.75%

Supplemental Table 3. Accumulative Pareto Values based on number of bets

Month	Top 20%	Top 1%
1	80.29%	18.63%
2	84.58%	19.57%
3	86.44%	20.32%
4	88.07%	21.48%
5	89.15%	22.51%
6	90.05%	22.94%
7	90.49%	23.04%
8	90.87%	23.19%
9	91.13%	23.37%
10	91.38%	23.52%
11	91.49%	23.60%

12 (annual)	91.75%	23.89%

Supplementary Table 4. Accumulative Pareto Values based on net loss

Number of months	Top 20% (primary analysis)	Net Loss users only	Top 20% (net gains included)	Top 1%
1	81.34%	81.34%	103.18%	23.75%
2	83.68%	83.68%	100.60%	27.60%
3	85.21%	85.21%	97.84%	28.56%
4	86.61%	86.60%	99.80%	29.16%
5	87.36%	87.36%	98.64%	29.52%
6	88.03%	88.03%	100.11%	29.71%
7	88.45%	88.45%	99.45%	30.20%
8	88.85%	88.85%	99.38%	30.17%
9	89.19%	89.19%	98.19%	30.45%
10	89.49%	89.49%	99.96%	30.30%
11	89.65%	89.65%	99.23%	30.37%

12 (annual)	89.87%	89.90%	98.67%	30.45%

Supplemental Table 5. Pareto Values based on Individual Months

Month	Number of Bets	Net Loss
October 2014	80.29%	81.34%
November 2014	81.78%	81.48%
December 2014	80.69%	81.45%
January 2015	81.95%	81.64%
February 2015	81.70%	81.03%
March 2015	81.29%	81.49%
April 2015	79.71%	80.58%
May 2015	79.71%	80.98%
June 2015	79.62%	80.85%
July 2015	80.00%	81.74%
August 2015	78.45%	80.79%

September 2015	79.85%	81.53%