Received: 7 March 2018 | Revised: 13 September 2018 | Accepted: 17 September 2018
DOI: 10.1002/mpr.1746

ORIGINAL ARTICLE

WILEY

Typology of online lotteries and scratch games gamblers' behaviours: A multilevel latent class cluster analysis applied to player account-based gambling data

Bastien Perrot[1,2,3] | Jean-Benoit Hardouin[1,3] | Marie Grall-Bronnec[1,2] | Gaëlle Challet-Bouju[1,2]

[1] Université de Nantes, Université de Tours, INSERM, SPHERE U1246 “methodS in Patient-centered outcomes and HEalth ResEarch”, Nantes, France

[2] Department of Addictology and Psychiatry, CHU Nantes, Nantes, France

[3] Biostatistics and Methodology Unit, Department of Clinical Research and Innovation, CHU Nantes, Nantes, France

Correspondence
Bastien Perrot, U1246 SPHERE - IRS2, 22 Boulevard Bénoni Goullin, Nantes 44200, France.
Email: bastien.perrot@univ-nantes.fr

Funding information
French Social Security Scheme for Liberal Professionals (RSI); French Inter-Departmental Agency for the Fight against Drugs and Addictive Behaviours (Mildeca); French National Institute of Health and Medical Research (INSERM); French National Institute for Prevention and Education in Health (INPES); French National Cancer Institute (INCA); ARC Foundation for Cancer Research; French Directorate General of Health; French National Health Insurance Fund for Employees (CNAMTS)

Abstract

Objectives: Internet gambling is often considered as a risk factor for gambling problems compared with land-based gambling. In parallel, this online activity generates data that can be useful to characterize Internet gambling behaviours. The objectives were to define a typology of online lotteries and scratch games gamblers' behaviours in order to identify early risky gambling behaviours and to classify gamblers in order to identify individuals with global risky gambling behaviours.

Methods: We performed a multilevel latent class cluster based on player account-based data of 10,000 gamblers from a French online operator.

Results: We identified seven clusters of online lotteries and scratch games gamblers' behaviours. A small cluster (3%) was characterized by a very high gambling activity, a high probability of chasing behaviour, a large proportion of bets concerning instant lotteries and scratch games, and a high proportion of women. We also found a group of gamblers having an 81% probability of being each month in this cluster.

Conclusions: The identification of distinct clusters of gambling behaviours and of groups of gamblers having different probabilities of being in these clusters through time could allow the implementation of personalized prevention measures according to the gamblers' profile.

KEYWORDS

gambling behaviour, latent class cluster analysis, online gambling, online lotteries, problem gambling

1 | INTRODUCTION

The Internet medium is often considered as a risk factor for gambling problem compared with land-based gambling. Indeed, gambling through an Internet-connected device, such as a computer or a smartphone, may give the user a feeling of immersion and anonymity, thus increasing the convenience of gambling activity (Griffiths, 2003). Online gambling is also characterized by a greater accessibility, more rapid and more frequent gambling outcomes, more available betting options, and the use of digital money (Gainsbury, 2015; Hing et al., 2015). These factors could contribute to increase the risk to develop gambling problems and may lead to disordered gambling. Gambling disorder is defined as a “persistent and recurrent problematic gambling behaviour leading to clinically significant impairment or distress” in the *Fifth Edition of the Diagnostic and Statistical Manual of Mental Disorders* (*DSM-5*; American Psychiatric Association, 2013).

Nevertheless, Internet gambling has the potential for generating a lot of information about the gambling activity of users. Player account-based gambling data refer to the records related to the personal gambling activity of a given operator client (Gainsbury, 2011). It includes, for example, the number of bets, the amount of money wagered, the number of money deposits, the use of loyalty bonuses, and the betting limits set by the user. Player account-based gambling data can be used to explore and characterize gambling behaviours and even give the possibility to screen for online gambling problems. For example, Adami et al. (2013), Braverman and Shaffer (2010), LaBrie and Shaffer (2011), and LaPlante, Schumann, LaBrie, and Shaffer (2008) used data from online sport gamblers; LaPlante, Kleschinsky, LaBrie, Nelson, and Shaffer (2009) and Luquiens et al. (2016) used data from online poker gamblers; Dragicevic, Tsogas, and Kudic (2011) and LaBrie, Kaplan, LaPlante, Nelson, and Shaffer (2008) used data from online casino gamblers. At the present time, such player account-based analysis of gambling behaviours does not exist for online lotteries and scratch games. Moreover, identifying specific gambling behaviours is important to apply effective responsible gambling measures and propose the most suitable intervention towards gamblers according to their characteristics (Caillon, Grall-Bronnec, Hardouin, Venisse, & Challet-Bouju, 2015; Cunningham, Hodgins, Toneatto, Rai, & Cordingley, 2009).

In France, the online gambling sector is regulated by the Regulatory Authority for Online Gambling (ARJEL) since 2010. The ARJEL issues licences to gambling operators for three types of online games: horse race betting, sports betting, and poker. Because of a particular waiver, only a single operator is allowed to provide online lotteries and scratch games. According to Costes, Eroukmanoff, Richard, and Tovar (2015), 52.7% of 15–75 years old French online gamblers played at least once to online lotteries in 2014 and 13.5% to online or scratch games. Online lotteries are by far the most played games on the Internet in France.

The approach generally used to identify distinct gambling behaviours is the identification of homogenous clusters of behaviours (or gamblers), that is, grouping similar observations together and then characterizing the clusters. Several clustering methods can be used to achieve this, the most often used being the *k*-means algorithm (Adami et al., 2013; Braverman & Shaffer, 2010; Dragicevic et al., 2011). However, the *k*-means method has several drawbacks such as the difficulty to handle sets of variables with different scale types, relatively strong assumptions on the distribution of variables (in particular normal assumption), and impossibility to make formal inference (Magidson & Vermunt, 2002; Steinley & Brusco, 2011). Thus, other clustering methods may be preferred. For example, latent class cluster analysis (LCCA; Vermunt & Magidson, 2002) is a model-based approach that assumes that the population is heterogeneous and that the data come from a mixture of cluster-specific distributions. The term "latent class" lies in the fact that the model hypothesizes the existence of a discrete latent variable that can take *K* values and that observations are assumed to belong to one of those *K* classes. Advantages of LCCA include the estimation of cluster membership probabilities (rather than exact affectations), the possibility to compute statistical criteria based on the likelihood of the model in order to choose the optimal number of clusters, and the possibility to handle mixed data (normal, binomial, count data, etc.).

In this study, we performed an LCCA to define a typology of online lottery and scratch games behaviours based on player account-based gambling data. Especially, we use a multilevel LCCA (Vermunt, 2008) to classify both monthly gambling behaviours and gamblers themselves. Indeed, in addition to identifying specific gambling behaviours, identifying specific groups of gamblers may also be of interest for the following reason: a given gambler A who had a risky gambling behaviour during 1 month and a "normal" gambling behaviour during other months should be seen as different from an another gambler B who had a risky gambling behaviour on almost every month.

2 | METHOD

The present work is part of the EDEIN study—first stage, previously described in Perrot et al. (2017; NCT02415296).

2.1 | Participants

The data used in this study are gambling-related records from a randomly selected sample of 10,000 clients of the French national lottery and scratch games operator. This operator is the only one in France who is allowed to propose lotteries and scratch games on the Internet. As a consequence, this operator represents 100% of the online lotteries and scratch games market in France, and the present sample can be considered as a very representative sample of these online games.

Only gamblers who played on the Internet at least once a lottery or scratch game between September 2015 and August 2016 were included.

This study was based on a retrospective gambling-related records analysis and was approved by the local research ethics committee Groupe Nantais d'Ethique dans le Domaine de la Santé (GNEDS) on March 25, 2015. The retrospective and non-interventional design of this study made the consent of the patients unnecessary.

Characteristics of gamblers and their gambling activity are described in Table 1.

TABLE 1 Description of gamblers and monthly gambling activity

	N = 10,000
Age (years)[a]	44 [19–92]
Male[b]	6819 (68.2%)
Age of account (months)[a]	46 [1–183]
Total stake per month (euros)[a]	12 [0–3918]
Number of bets per month[a]	2 [0–1351]
Total deposit per month (euros)[a]	9 [0–1749]
Biggest deposit in a single day (euros)[a]	8 [0–534]
Number of gambling days per month[a]	1 [0–30]
Number of different games played[a]	1 [0–44]
Use of bonuses (frequency)[c]	6.1% ± 13.3%
Chasing proxy (frequency)[c]	6.5% ± 14.8%
Percentage of delayed lottery bets[c]	50.0% ± 34.8%

Note. Gambling indicators were averaged for each gambler over the 12 months before computing the descriptive statistics. Consequently, the numbers in this table are based on the monthly average activity of each gambler.

[a]Median [range].

[b]Frequency (%).

[c]Mean ± standard deviation.

2.2 | Data

We chose the month level to describe gambling behaviours, so that we were able to observe 12 gambling behaviours (i.e., one behaviour per month between September 2015 and August 2016) for each gambler. Available variables were age of the gambler (years), gender, age of the account (months), total money wagered per month (euros), number of bets per month, total deposit per month (euros), biggest deposit in a single day (euros), number of gambling days per month,[1] number of different games played, use of loyalty bonuses (yes/no), chasing proxy (yes/no),[2] and percentage of delayed lottery bets.[3]

2.3 | Indicators of "unusual" intragambler activity

For the LCCA, we included the variables described above (excluding age of gambler, gender, age of account, and percentage of delayed lottery bets, which were included as inactive covariates[4]). We also included in addition two variables related to the evolution of gambling activity over time. The objective in including these variables was to detect an unusual intragambler variation of gambling behaviour, which might be associated to a loss of control over the gambling practice.

The first evolution indicator measures the changes in the number of gambling days over the studied months. Specifically, we regress the number of gambling days at month m on the mean of the number of gambling days computed on the months $m - 1$, $m - 2$, and $m - 3$ (the "baseline value"). We added a random effect on the baseline value in order to take into account the heteroscedasticity. Once the regression parameters estimated, we computed for each observation the ratio between the residual (observed value-predicted value) and the square root of the total variance of the prediction. Then the ratios were standardized (mean = 0, variance = 1) in order to be more easily interpretable (e.g., ratios >2 could be considered "extreme"). We applied the same method for the changes in total money wagered. The rationale of this method is that large residuals correspond to values that are not explained by the baseline value and thus can be considered as unexpected/unusual gambling activity.

2.4 | Methodology and statistical analysis

We used the methodology described in the study protocol by Perrot et al. (2017), that is to say we performed a clustering of monthly gambling behaviours where the statistical unit is a person-month and simultaneously a clustering of gamblers where the statistical unit is a person. Gamblers can thus be affected to different person-month clusters from a month to another.

2.4.1 | Multilevel latent class cluster analysis

LCCA is a probabilistic model, which assumes that the population of interest is heterogeneous and composed of several homogeneous subpopulations called clusters. Each cluster is defined by a specific (multivariate) probability distribution; the mixture of these cluster-specific distributions forms the marginal probability distribution, which describes the population. Once the parameters (the size of the clusters and the parameters of the cluster-specific distributions) of the mixture model are estimated, it is possible to compute the probability that an observation belong to each cluster. In the mixture model, the clusters correspond to the values taken by a latent discrete variable (e.g., a latent variable whose possible values indicate a specific gambling behaviour). In our context, for each month, each gambler has a set of probabilities to belong to each cluster. For example, assuming there are three clusters for a given month, a gambler could have a 0.70 probability to belong to the first cluster, 0.20 to belong the second cluster, and 0.10 to belong to the third cluster. The next month, the respective probabilities could be 0.60, 0.30, and 0.10, and so on.

Additionally, by using a multilevel LCCA (Vermunt, 2003), we can classify the gamblers based on the distribution of their monthly cluster-membership probabilities. For the sake of clarity, we will use the following terms: "clusters" represent the groups of monthly gambling behaviours (person-month, first level units) and "classes" represent the groups of gamblers (person, second level units). With the previous three-cluster example, the concerned gambler could be classified in a class of gamblers characterized by a high probability to be in the first cluster and a low probability to be in the third cluster. We used a parameterization of multilevel LCCA where class-membership probabilities are not directly influenced by gambling indicators but rather depend on the distribution of the cluster-membership probabilities; only cluster-membership probabilities are influenced by monthly gambling indicators (Vermunt, 2008). This approach offers the dual advantage of taking into account the non-independence of observations assumed by LCCA (because of the nested structure of the data) and getting an insight of evolutions of gambling behaviours over months for a specific gambler.

2.4.2 | Conditional cluster distributions

The conditional distribution forms were chosen depending on the scale type of each indicator. Log-normal distribution was assumed for total money wagered, total deposit, and biggest deposit in a single day; normal distribution was assumed for the number of bets; overdispersed Poisson distribution was assumed for number of gambling days per month and number of different games played; and binomial distribution was assumed for use of loyalty bonus and chasing proxy. Normal distribution was assumed for the indicator of "unusual" number of gambling days and the indicator of "unusual" total money wagered.

[1]A gambling day is defined as a day during which the gambler has gambled at least once.

[2]The chasing proxy is a binary indicator defined as follows: It equals 1 if we observed a sequence of three money deposits within a 12-hr period or if we observe a deposit made less than 1 hr after a bet. Otherwise, it equals 0.

[3]Delayed lotteries are lottery games whose result is known after a significant delay. There are opposed to instant lotteries whose result is known almost immediately after gambling. The third type of game is scratch games, which are considered similar to instant lotteries in this context.

[4]Total deposit, biggest deposit in a single day, and total money wagered were also added as inactive covariates for interpretation purpose because they were log-transformed in the analysis. The means of covariates in each cluster were computed as weighted means, taking into account the individual cluster membership probabilities.

2.4.3 | Estimation

Estimation of the models was made using Latent Gold 5.1 (Vermunt & Magidson, 2015). Parameters were estimated using maximum a posteriori with negligible informative prior, which yields point estimates similar to those obtained with maximum likelihood while avoiding boundary estimates (Garre & Vermunt, 2006).

2.4.4 | Model selection

We fit multiple models in order to determine the most appropriate number of clusters and classes. First, we fit one-class models composed of one to 10 clusters. Once the number of clusters was selected, the corresponding model was tested with one to six classes. A general rule of thumb is to select the model minimizing the negative log-likelihood function augmented by a penalty function that increases with the complexity of the model (i.e., the number of parameters and/or the number of observations). The Bayesian information criterion (BIC) is often used to this end. In this study, we performed the model selection based on a trade-off between the BIC, the classification error rate (which reflects the precision of the classification), and the interpretability of the clusters. To prevent ending up with a local solution, we used multiple sets of random starting values.

3 | RESULTS

3.1 | Model selection

Fit indices of one- to 10-cluster models are given in Table 2. Based solely on the BIC criteria, the 10-cluster model would be selected. However, as noted by Thøgersen (2017) and van den Bergh et al. (2017), with large data sets, the BIC has a tendency to keep decreasing as the number of clusters increases. Classification error rates were very good for all tested solutions. We chose the seven-cluster solution because models with more than seven clusters yield to additional clusters characterized by either nonmeaningful interpretation or which seemed very close to other clusters. For example, the eight-cluster model was similar to the seven-cluster solution but with a very small cluster (1%) characterized by no money wagered but some money was still deposited on the account. With a closer look at the data, this rare behaviour corresponds to people who either make deposits of money in month *m* and use this money to gamble in month *m* + 1 (this behaviour can make sense) or people who make deposits of money in month *m* but do not gamble in month *m*, in month *m + 1*, or in month *m + 2*. This, however, makes less sense or at least is not very interesting and concerns very few observations.

Table 3 shows the fit indices for one- to seven-class seven-cluster models. Similarly, we did not observe neither a minimum BIC for the tested models or a significant difference in the classification error rates. Because the seven-class solution did not bring meaningful information, we selected the six-class solution, which in addition (comparing with the five-class solution) allowed observing a class of gamblers having a high probability of being in the cluster seven every month.

TABLE 2 Fit indices of one-level latent class cluster analysis

	Log-likelihood	BIC	Classification errors
1-Cluster	−1,583,047	3,166,299	0.000
2-Cluster	−686,860	1,374,118	0.001
3-Cluster	−354,053	708,697	0.001
4-Cluster	−190,415	381,615	0.007
5-Cluster	−138,138	277,253	0.022
6-Cluster	−108,257	217,684	0.032
7-Cluster	−79,677	160,718	0.034
8-Cluster	−59,335	120,227	0.017
9-Cluster	−35,028	71,805	0.032
10-Cluster	−18,347	38,638	0.031

Note. BIC: Bayesian information criterion.

3.2 | Parameter estimates

The higher part of Table 4 shows the distribution of the seven monthly gambling behaviours clusters across the six classes of gamblers. The middle part shows the estimated means of the indicators of the cluster -distributions. The lower part shows the parameter estimates of inactive covariates.

3.3 | Description of clusters of monthly gambling behaviours

Cluster 1 (40% of the monthly gambling behaviours) corresponded to gambling inactivity, meaning no money was wagered or deposited on the gambler's account during the month.

Cluster 2 (13% of the monthly gambling behaviours) was characterized by a low gambling activity (in average: 11 euros wagered, 1.8 gambling days, and no deposit of money).

Cluster 3 (14% of the monthly gambling behaviours) was similar to Cluster 2, with slightly higher amounts of money wagered (22 euros) and deposited on the account (25 euros).

Cluster 4 (4% of the monthly gambling behaviours) was characterized by a higher gambling activity compared with Clusters 2 and 4 (34 euros wagered, 5.6 gambling days, and 11 bets) but with no deposits of money.

Cluster 5 (14% of the monthly gambling behaviours) was close to Cluster 4 in terms of money wagered (30 euros) and number of gambling days (5.1). Deposits of money were similar to those observed in Cluster 3 (25 euros), and the probability of chasing was 9%. The indicator measuring the “deviation” from the usual number of gambling days was 0.54, close to the value for Cluster 6.

TABLE 3 Fit indices of two-level latent class cluster analysis

	Log-likelihood	BIC	Classification errors
7-Cluster 1-Class	−79,677	160,718	0.034
7-Cluster 2-Class	−61,505	124,453	0.030
7-Cluster 3-Class	−53,310	108,143	0.031
7-Cluster 4-Class	−49,312	100,226	0.026
7-Cluster 5-Class	−46,751	95,184	0.018
7-Cluster 6-Class	−46,066	93,892	0.025
7-Cluster 7-Class	−44,947	91,735	0.025

Note. BIC: Bayesian information criterion.

TABLE 4 Cluster sizes, class sizes, and cluster-specific parameter estimates of the 7-cluster 6-class model

	Class size	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7
Class 1	0.39	0.77	0.11	0.07	0.01	0.03	0.01	0.00
Class 2	0.27	0.20	0.28	0.38	0.02	0.11	0.02	0.00
Class 3	0.12	0.02	0.10	0.04	0.08	0.69	0.07	0.00
Class 4	0.09	0.37	0.05	0.06	0.13	0.10	0.24	0.06
Class 5	0.09	0.00	0.00	0.01	0.09	0.04	0.81	0.04
Class 6	0.03	0.01	0.00	0.00	0.05	0.00	0.13	0.81
Average cluster size		0.40	0.13	0.14	0.04	0.14	0.12	0.03
Estimated means of the indicators								
Total deposit (log-euros)		0.00	0.00	3.05	0.00	3.18	3.99	4.94
Biggest deposit in a single day (log-euros)		0.00	0.00	2.99	0.00	2.93	3.41	3.81
Chasing (yes)		0.00	0.00	0.04	0.02	0.09	0.20	0.61
Total wager (log-euros)		0.00	2.22	2.78	3.00	3.34	4.32	5.69
Number of bets		0.00	2.34	1.98	11.20	7.66	19.28	207.61
Number of gambling days		0.00	1.83	1.58	5.55	5.13	7.98	12.46
Number of different games played		0.00	1.34	1.30	2.25	2.06	3.23	10.82
Use of loyalty bonuses (yes)		0.00	0.03	0.03	0.12	0.06	0.15	0.52
"Deviation" from usual total money wagered		−0.25	−0.10	0.03	−0.27	0.10	0.37	2.07
"Deviation" from usual number of gambling days		−0.46	0.01	0.01	0.22	0.54	0.58	1.11
Inactive covariates								
Percentage of deferred lottery bets (%)		0.00	93.15	92.44	65.03	87.26	70.88	18.53
Age of gambler (years)		43.15	46.46	45.30	47.68	47.49	49.74	50.29
Age of account (months)		54.00	61.34	60.07	60.55	62.09	68.29	60.16
Male		0.66	0.70	0.72	0.72	0.70	0.72	0.54
Total deposit (euros)		0.00	0.00	25.17	0.00	25.41	65.09	240.89
Biggest deposit in a single day (euros)		0.00	0.00	23.07	0.00	19.57	36.17	66.27
Total money wagered (euros)		0.00	10.86	21.84	33.89	30.15	87.45	482.83

Cluster 6 (12% of the monthly gambling behaviours) was characterized by a higher gambling activity compared with the first five clusters (87 euros wagered, eight gambling days, 65 euros deposited, and 19 bets per month). The two indicators of "unusual gambling activity" were 0.37 for the total money wagered and 0.58 for the number of gambling days, indicating a trend for an "unusual" intragambler activity. The probability of observing a chasing behaviour was 20%.

Cluster 7 (3% of the monthly gambling behaviours) was characterized by a very high gambling activity compared with other clusters (483 euros wagered, 12.5 gambling days, 241 euros deposited, and 208 bets per month). The two indicators of "unusual gambling activity" were 2.07 for the total money wagered and 1.11 for the number of gambling days. The proportion of chasing behaviour was 61%. The diversity of games played was more important (12 different games played) than in other clusters, as well as the probability of using loyalty bonuses (52%). The percentage of delayed lottery bets was only 19%, which means that most of the bets concerned scratch games or instant lotteries. Finally, this was the cluster with the highest percentage of women (46%).

3.4 | Description of classes of gamblers

Gamblers in Class 1 (39% of the gamblers) had 77%, 11%, and 7% probabilities of being each month in Cluster 1, 2, and 3, respectively. That means 39% of individuals did not gamble most of the time and gambled sparsely throughout the year.

Class 2 (27% of the gamblers) was composed of gamblers who had 38%, 28%, and 20% probabilities of being in Clusters 3, 2, and 1, respectively. They were likely to be occasional gamblers who either did not gamble during a whole month (Cluster 1) or wage small amounts of money (Clusters 2 and 3).

Gamblers in Class 3 (12% of the gamblers) had a 69% probability of belonging to Cluster 5 and low probabilities (from 0% to 10%) of being in other clusters. As a consequence, they were likely to have a "moderate" gambling activity quite stable over time.

Class 4 (9% of the gamblers) was composed of gamblers who had 37%, 24%, and 13% probabilities of being in Clusters 1, 6, and 4, respectively. These are probably individuals who transited from no gambling periods (Cluster 1) to high gambling activity periods (Cluster 5) over time, and vice versa.

Gamblers of Class 5 (9% of the gamblers) had an 81% probability of being in Cluster 6; thus, they seemed to have a stable high gambling activity over time.

Class 6 (3% of the gamblers) was characterized by an 81% probability of belonging to Cluster 7 and a 13% probability of belonging to Cluster 6. They could correspond to gamblers with a risky gambling behaviour, that is, with a lot of bets, many different games played, a high probability of chasing, and a substantial risk of having an "unusual" gambling activity observed during a month. The proportion of women in this class was the highest (48%).

4 | DISCUSSION

4.1 | Main findings

The goal of this study was to define a typology of online gambling behaviours and of online gamblers based on player account-based gambling data. In the context of online lotteries and scratch games, a multilevel latent class cluster analysis revealed seven clusters of monthly gambling behaviours that differed with respect to gambling indicators and six classes of gamblers that differed with respect to the clusters' membership probabilities.

The most prevalent cluster (Cluster 1, 40%) corresponded to gambling inactivity. Clusters 2, 3, 4, and 5 were characterized by a low to moderate gambling activity. Nevertheless, the probability of chasing behaviour in Cluster 5 was 9%. Cluster 6 was characterized by a higher gambling activity that might be seen as risky; the 20% probability of observing a chasing behaviour suggests a relatively moderate proportion of loss of control in this cluster, compared with the other clusters. A small cluster (Cluster 7, 3%) of gambling behaviours was characterized by a very high gambling activity associated with a high proportion of chasing, large “deviations” from their usual gambling activity, and a high level of involvement measured by the number of different games played. Breadth involvement has been found significantly associated with gambling disorder (LaPlante, Nelson, & Gray, 2014). In this cluster, most of the bets concerned lotteries with instant results and scratch games, even though this indicator was included as an inactive covariate in the model. These types of games are often associated with gambling problems more than delayed lotteries especially because of a high event frequency (Challet-Bouju et al., 2015; Husky, Michel, Richard, Guignard, & Beck, 2015; McCormack & Griffiths, 2013). For these reasons, we could hypothesize that Cluster 7 corresponded to risky monthly gambling behaviours. A small class (3%) of gamblers had an 81% probability of being in this particular cluster, so that we can assume that they represent gamblers at risk for gambling problems. This proportion is very close to the prevalence of excessive or at moderate risk gamblers in the last national French prevalence survey within gamblers who played preferentially to lotteries and scratch games (3%; Costes et al., 2015). Moreover, gamblers in Classes 7 had very low probabilities of being in Clusters 1, 2, and 3, which means that they had no significant period of gambling abstinence. This is a further indicator suggesting the “at-risk” status of gamblers of Class 7 because Bruneau et al. (2016) found that experiencing at least 1 month of gambling abstinence was a protective factor against transition to problem gambling.

Our sample was composed of a high proportion of women (32%), compared with other studies on Internet gambling (LaPlante et al., 2009: 5.5%; LaPlante et al., 2014: 5.3%; Braverman & Shaffer, 2010: 8%; Luquiens et al., 2016: 10%). This high proportion is not surprising given the specificity of our sample, which includes only gamblers of online lotteries and scratch games. Indeed, it is well known that women tend to prefer pure chance games (lotteries and scratch games) over other types of games such as sports betting, horse race betting, or poker (McCormack, Shorter, & Griffiths, 2014). However, a very interesting finding in the multilevel LCCA is that the proportion of women is even much higher in Cluster 7 (46%; 48% of gamblers in Class 6). At first sight, the high proportion of women in this cluster may appear surprising because men are usually often found more at risk of gambling problems than women (Hing et al., 2015; Williams, Volberg, & Stevens, 2012). However, Cluster 7 is associated to a high proportion of instant lotteries and scratch games compared with delayed lotteries. In a recent study focusing on-site lottery and scratch games gamblers, Brochado, Santos, Oliveira, and Esperança (2018) showed that women were more likely to engage in scratch games and that men preferred delayed lotteries. This is consistent with the specificity of Cluster 7. Moreover, one can assume that women gamblers who encounter gambling problems might gamble preferentially on the Internet to take advantage of anonymity and gamble without the feeling of stigma (Griffiths, 2003). It has also been suggested that women problem gamblers are more socially anxious and avoid social activities compared with either male problem gamblers or women in general (Corney & Davis, 2010). Because online gambling allows playing in isolation, this may be an attractive option for women problem gamblers.

The fact that online instant lotteries and scratch games seem related to a risky gambling behaviour allows considering the possibility to propose targeted prevention measures with respect to gamblers' favourite game. For example, we could integrate information messages or money/time limitation for instant lotteries or scratch games gamblers who make multiple deposits of money during their gambling session. The presence of a probable chasing behaviour and of a high level of breadth involvement (in Cluster 7) suggests a loss of control over the gambling behaviour that could be tackled by displaying messages related to chasing (e.g., when a gambler deposits money on his or her account following a loss, a message could suggest having a break before investing money again).

Moreover, we observed that individuals with a risky gambling behaviour (Cluster 7) tended to have an “unusual” gambling activity. A way to improve the early detection of at-risk gamblers may be to take into account the changes in gambling activity as significantly as the raw gambling activity. A live screening algorithm that combines, individually for each gambler, both raw gambling activity and changes over time may represent a very interesting advance in the prevention of gambling problems online. This is the subject of the EDEIN study (Perrot et al., 2017; NCT02415296).

4.2 | Limitations

Some limitations must be noted in this study. Regarding LCCA itself, the choice of the “best” model can be influenced by misspecification of cluster-specific distributions. In particular, the BIC may suggest a more complex model (i.e., a model with more clusters) when an inadequate distributional form is assumed for a given variable (Bauer & Curran, 2004). This is typically the case when we assume normal distributions for highly skewed variables, such as total money wagered or total deposit. Even with a log-transformation, extreme values may lead to overextraction of clusters (Bauer & Curran, 2003). Large samples can also make it difficult to choose the number of clusters based on the BIC (Paas, 2014; Thøgersen, 2017). Furthermore, several significant bivariate residuals indicated that the local independence assumption was not satisfied. This could have affected the results by giving too much weight to some indicators. We tried to increase the number of clusters but the problem remained present, and this strategy led to the disadvantage of having clusters without meaningful interpretation. We also attempted to relax the local independence assumption for pairs of indicators (e.g., by starting with the largest bivariate residual). When doing this, the estimation procedure did not converge.

Regarding the classification of gambling behaviours, interpretation about gambling problems should be taken with caution. Although we tried using pertinent behavioural factors presumably associated with a risk of gambling problems, available data are not sufficient to clearly affirm that we found a group of risky gambling behaviours. As noted by Gainsbury (2011), account-based data can be useful to identify risk factors, but there is a lack of contextual factors (level of income, using of multiple accounts, state of mind while gambling, consequences of gambling, etc.). In this context, Perrot et al. (2017) proposed to link account-based data of gamblers who would also answer a self-report questionnaire about gambling problems (the Problematic Gambling Severity Index questionnaire; Ferris & Wynne, 2001) and participate to clinical interviews based on the gambling disorder section of the *DSM-5*. Without these additional data, it is for now difficult to assert whether the risky gambling behaviours we observed take place before they become out of control (for problem gamblers) or rather are the consequences of a loss of control that would have already occurred. This implies wondering if the concerned gamblers have entered in the “vicious circle” of pathological gambling or not (Sharpe, 2002). In this context, the use of the chasing proxies, as defined in this study, may be of interest to identify early risky gambling behaviours. Indeed, the chasing behaviour is closely related to the loss of control and is thought to represent the last step before pathological gambling (Blaszczynski & Nower, 2002). In the future, it will be important to be able to make this distinction in the operationalization of prevention measures, in particular to adapt messages sent to the gamblers.

Regarding the indicators of “unusual” gambling activity, the proposed method based on the residuals of the regression of current values on “baseline” values does not provide an easy interpretation of the deviations from “usual” gambling activity.

Despite these limitations, this is to our knowledge the first study to focus on gambling behaviours of online lotteries and scratch games gamblers, in a highly representative and large sample. It allowed emphasizing a high proportion of women in a group of presumably problem gamblers. Moreover, the use of an original method (multilevel LCCA) had allowed avoiding certain problems encountered with more traditional statistical approaches. In particular, it allows differentiating atypical behaviours, even with small groups (other methods often gather atypical observations into a single group, leading to a loss of information). Differentiated small atypical groups may precisely correspond to the expected at-risk groups. Finally, the inclusion of indicators related to the variability of the gambling activity over time prevents to detect simply “big” players and allows identifying probable problem gamblers. This kind of indicator can also allow for a very early detection of at-risk gamblers and thus may allow setting up preventive actions very quickly.

The use of player account-based gambling data seems promising for the identification of specific online gambling behaviours. It could allow setting up personalized prevention measures based on particular gambling behaviours. Results from multilevel LCCA suggest that some gamblers may transit from a gambling behaviour to another over time. It would thus be interesting to study these changes in the gamblers' status over time more specifically to identify the behavioural factors associated with these transitions.

ACKNOWLEDGEMENTS

We would like to warmly thank the FDJ for their contribution to this project, by allowing us to have access to the gamblers' data. This research was conducted on the initiative of and coordinated by the Department of Addictology and Psychiatry of the University Hospital of Nantes, which sponsors this study.

DECLARATION OF INTEREST STATEMENT

M.G.-B. and G.C.-B. declare that the University Hospital of Nantes has received gambling industry (FDJ and PMU) funding in the form of a sponsorship that supports the gambling section of the BALANCED Unit (the Reference Centre for Excessive Gambling). Scientific independence towards gambling industry operators is warranted. There were no publishing constraints. Other authors declare that they have no conflict of interest.

FUNDING

This research has benefited from the joint assistance of the French National Health Insurance Fund for Employees (CNAMTS), the French Directorate General of Health, the ARC Foundation for Cancer Research, the French National Cancer Institute (INCA), the French National Institute for Prevention and Education in Health (INPES), the French National Institute of Health and Medical Research (INSERM), the French Inter-Departmental Agency for the Fight against Drugs and Addictive Behaviours (Mildeca), and the French Social Security Scheme for Liberal Professionals (RSI) as part of the “Primary Prevention” call for proposals issued by the French Institute for Public Health Research (IReSP) and INCA in 2013.

ORCID

Bastien Perrot http://orcid.org/0000-0002-3701-6693
Gaëlle Challet-Bouju http://orcid.org/0000-0002-2238-8005

REFERENCES

Adami, N., Benini, S., Boschetti, A., Canini, L., Maione, F., & Temporin, M. (2013). Markers of unsustainable gambling for early detection of at-risk online gamblers. *International Gambling Studies*, *13*(2), 188–204. https://doi.org/10.1080/14459795.2012.754919

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders (DSM-5®)*American Psychiatric Pub. https://doi.org/10.1176/appi.books.9780890425596

Bauer, D. J., & Curran, P. J. (2003). Distributional assumptions of growth mixture models: Implications for overextraction of latent trajectory classes. *Psychological Methods*, *8*(3), 338–363. https://doi.org/10.1037/1082-989X.8.3.338

Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods*, *9*(1), 3–29. https://doi.org/10.1037/1082-989X.9.1.3

Blaszczynski, A., & Nower, L. (2002). A pathways model of problem and pathological gambling. *Addiction (Abingdon, England)*, *97*(5), 487–499. https://doi.org/10.1046/j.1360-0443.2002.00015.x

Braverman, J., & Shaffer, H. J. (2010). How do gamblers start gambling: Identifying behavioural markers for high-risk internet gambling. *The European Journal of Public Health*. ckp232. https://doi.org/10.1093/eurpub/ckp232

Brochado, A., Santos, M., Oliveira, F., & Esperança, J. (2018). Gambling behavior: Instant versus traditional lotteries. *Journal of Business Research* https://doi.org/10.1016/j.jbusres.2018.01.001, *88*, 560–567.

Bruneau, M., Grall-Bronnec, M., Vénisse, J.-L., Romo, L., Valleur, M., Magalon, D., … Hardouin, J.-B. (2016). Gambling transitions among adult gamblers: A multi-state model using a Markovian approach applied to the JEU cohort. *Addictive Behaviors*, *57*, 13–20. https://doi.org/10.1016/j.addbeh.2016.01.010

Caillon, J., Grall-Bronnec, M., Hardouin, J.-B., Venisse, J.-L., & Challet-Bouju, G. (2015). Online gambling's moderators: how effective? Study protocol for a randomized controlled trial. *BMC Public Health*, *15*(1), 519. https://doi.org/10.1186/s12889-015-1846-7

Challet-Bouju, G., Hardouin, J.-B., Renard, N., Legauffre, C., Valleur, M., Magalon, D., … Grall-Bronnec, M. (2015). A gamblers clustering based on their favorite gambling activity. *Journal of Gambling Studies*, *31*(4), 1767–1788. https://doi.org/10.1007/s10899-014-9496-8

Corney, R., & Davis, J. (2010). The attractions and risks of Internet gambling for women: A qualitative study. *Journal of Gambling Issues*, *0*(24), 121–139. https://doi.org/10.4309/jgi.2010.24.8

Costes, J.-M., Eroukmanoff, V., Richard, J.-B., & Tovar, M.-L. (2015). Les jeux d'argent et de hasard en France en 2014. *Notes de l'ODJ n°6*, (6). Consulté à l'adresse http://www.academia.edu/12112626/Les_jeux_dargent_et_de_hasard_en_France_en_2014

Cunningham, J. A., Hodgins, D. C., Toneatto, T., Rai, A., & Cordingley, J. (2009). Pilot study of a personalized feedback intervention for problem gamblers. *Behavior Therapy*, *40*(3), 219–224. https://doi.org/10.1016/j.beth.2008.06.005

Dragicevic, S., Tsogas, G., & Kudic, A. (2011). Analysis of casino online gambling data in relation to behavioural risk markers for high-risk gambling and player protection. *International Gambling Studies*, *11*(3), 377–391. https://doi.org/10.1080/14459795.2011.629204

Ferris, J., & Wynne, H. (2001). *The Canadian problem gambling index*. Ottawa, ON: Canadian Centre on substance abuse.

Garre, F. G., & Vermunt, J. K. (2006). Avoiding boundary estimates in latent class analysis by Bayesian posterior mode estimation. *Behaviormetrika*, *33*(1), 43–59.

Gainsbury, S. (2011). Player account-based gambling: Potentials for behaviour-based research methodologies. *International Gambling Studies*, *11*(2), 153–171. https://doi.org/10.1080/14459795.2011.571217

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

Griffiths, M. (2003). Internet gambling: Issues, concerns, and recommendations. *Cyberpsychology & Behavior*, *6*(6), 557–568. https://doi.org/10.1089/109493103322725333

Hing, N., Cherney, L., Gainsbury, S. M., Lubman, D. I., Wood, R. T., & Blaszczynski, A. (2015). Maintaining and losing control during Internet gambling: A qualitative study of gamblers' experiences. *New Media & Society*, *17*(7), 1075–1095. https://doi.org/10.1177/1461444814521140

Husky, M. M., Michel, G., Richard, J.-B., Guignard, R., & Beck, F. (2015). Gender differences in the associations of gambling activities and suicidal behaviors with problem gambling in a nationally representative French sample. *Addictive Behaviors*, *45*, 45–50. https://doi.org/10.1016/j.addbeh.2015.01.011

LaBrie, R., & Shaffer, H. J. (2011). Identifying behavioral markers of disordered Internet sports gambling. *Addiction Research & Theory*, *19*(1), 56–65. https://doi.org/10.3109/16066359.2010.512106

LaBrie, R. A., Kaplan, S. A., LaPlante, D. A., Nelson, S. E., & Shaffer, H. J. (2008). Inside the virtual casino: A prospective longitudinal study of actual Internet casino gambling. *European Journal of Public Health*, *18*(4), 410–416. https://doi.org/10.1093/eurpub/ckn021

LaPlante, D. A., Kleschinsky, J. H., LaBrie, R. A., Nelson, S. E., & Shaffer, H. J. (2009). Sitting at the virtual poker table: A prospective epidemiological study of actual Internet poker gambling behavior. *Computers in Human Behavior*, *25*(3), 711–717. https://doi.org/10.1016/j.chb.2008.12.027

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: Journal of the Society of Psychologists in Addictive Behaviors*, *28*(2), 396–403. https://doi.org/10.1037/a0033810

LaPlante, D. A., Schumann, A., LaBrie, R. A., & Shaffer, H. J. (2008). Population trends in Internet sports gambling. *Computers in Human Behavior*, *24*(5), 2399–2414. https://doi.org/10.1016/j.chb.2008.02.015

Luquiens, A., Tanguy, M.-L., Benyamina, A., Lagadec, M., Aubin, H.-J., & Reynaud, M. (2016). Tracking online poker problem gamblers with player account-based gambling data only. *International Journal of Methods in Psychiatric Research*, *25*(4), 333–342. https://doi.org/10.1002/mpr.1510

Magidson, J., & Vermunt, J. (2002). Latent class models for clustering: A comparison with K-means. *Canadian Journal of Marketing Research*, *20*(1), 36–43.

McCormack, A., & Griffiths, M. D. (2013). A scoping study of the structural and situational characteristics of Internet gambling. *International Journal of Cyber Behavior, Psychology and Learning (IJCBPL)*, *3*(1), 29–49. https://doi.org/10.4018/ijcbpl.2013010104

McCormack, A., Shorter, G. W., & Griffiths, M. D. (2014). An empirical study of gender differences in online gambling. *Journal of Gambling Studies*, *30*(1), 71–88. https://doi.org/10.1007/s10899-012-9341-x

Paas, L. J. (2014). Comments on: Latent Markov models: A review of a general framework for the analysis of longitudinal data with covariates. *Test*, *23*(3), 473–477. https://doi.org/10.1007/s11749-014-0387-1

Perrot, B., Hardouin, J.-B., Costes, J.-M., Caillon, J., Grall-Bronnec, M., & Challet-Bouju, G. (2017). Study protocol for a transversal study to develop a screening model for excessive gambling behaviours on a representative sample of users of French authorised gambling websites. *BMJ Open*, *7*(5), e014600. https://doi.org/10.1136/bmjopen-2016-014600

Sharpe, L. (2002). A reformulated cognitive–behavioral model of problem gambling: A biopsychosocial perspective. *Clinical Psychology Review*, *22*(1), 1–25. https://doi.org/10.1016/S0272-7358(00)00087-8

Steinley, D., & Brusco, M. J. (2011). Evaluating mixture modeling for clustering: Recommendations and cautions. *Psychological Methods*, *16*(1), 63–79. https://doi.org/10.1037/a0022673

Thøgersen, J. (2017). Housing-related lifestyle and energy saving: A multilevel approach. *Energy Policy*, *102*, 73–87. https://doi.org/10.1016/j.enpol.2016.12.015

van den Bergh, M., Schmittmann, V. D., & Vermunt, J. K. (2017). Building latent class trees, with an application to a study of social capital. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 13(Supplement 1), 13–22. https://doi.org/10.1027/1614-2241/a000128

Vermunt, J. K. (2008). Latent class and finite mixture models for multilevel data sets. *Statistical Methods in Medical Research*, *17*(1), 33–51. https://doi.org/10.1177/0962280207081238

Vermunt, J. K. (2003). Multilevel latent class models. *Sociological Methodology*, *33*(1), 213–239. https://doi.org/10.1111/j.0081-1750.2003.t01-1-00131.x

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. *Applied Latent Class Analysis*, 89–106. https://doi.org/10.1017/CBO9780511499531.004

Vermunt, J. K., & Magidson, J. (2015). *Technical guide for latent gold 5.1: Basic, advanced, and syntax*. Belmont, MA: Statistical Innovations Inc.

Williams, R. J., Volberg, R. A., & Stevens, R. M. (2012). *The population prevalence of problem gambling: Methodological influences, standardized rates, jurisdictional differences, and worldwide trends*. Ontario Problem Gambling Research Centre. Consulté à l'adresse https://www.uleth.ca/dspace/handle/10133/3068

How to cite this article: Perrot B, Hardouin J-B, Grall-Bronnec M, Challet-Bouju G. Typology of online lotteries and scratch games gamblers' behaviours: A multilevel latent class cluster analysis applied to player account-based gambling data. *Int J Methods Psychiatr Res*. 2018;27:e1746. https://doi.org/10.1002/mpr.1746