UNIVERSITE BIOLOGIE BRETAGNE SANTE LOIRE

UNIVERSITÉ DE NANTES

THESE DE DOCTORAT DE

L'UNIVERSITE DE NANTES
COMUE UNIVERSITE BRETAGNE LOIRE

ECOLE DOCTORALE N° 605
Biologie Santé
Spécialité : *Epidémiologie, Analyse de Risque, Recherche Clinique*

Par
Bastien PERROT

Repérage des problèmes liés à la pratique des jeux de hasard et d'argent sur Internet

Thèse présentée et soutenue à Nantes, le 27 novembre 2018
Unité de recherche : UMR INSERM 1246 SPHERE « methodS in Patient-centered outcomes and HEalth ResEarch »

Rapporteurs avant soutenance :

Amine BENYAMINA — Professeur des universités - Praticien hospitalier, Université Paris-XI
Cécile PROUST-LIMA — Chargée de recherche, Université de Bordeaux

Composition du Jury :

Président : Bruno FALISSARD	Professeur des universités - Praticien hospitalier, Université Paris-XI
Examinateur : Bruno FALISSARD	Professeur des universités - Praticien hospitalier, Université Paris-XI
Dir. de thèse : Jean-Benoit HARDOUIN	Maître de conférences des universités, Université de Nantes
Co-dir. de thèse : Marie GRALL-BRONNEC	Professeur des universités - Praticien hospitalier, Université de Nantes
Co-enc. de thèse : Gaëlle CHALLET-BOUJU	Ingénieur de recherche, Université de Nantes

Valorisations scientifiques

Publications issues de la thèse

- 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. doi.org/10.1136/bmjopen-2016-014600
- Perrot, B., Hardouin, J.-B., Grall-Bronnec, M., Challet-Bouju, G. (2018). Typology of online lotteries and scratch games gamblers' behaviours : a multilevel latent class cluster analysis applied to player account-based gambling data. (accepté pour publication dans *International Journal of Methods in Psychiatric Research*)

Autres publications en lien avec la thèse

- Caillon, J., Grall-Bronnec, M., Perrot, B., Leboucher, J., Donnio, Y., Romo, L., & Challet-Bouju, G. (2018). Effectiveness of At-Risk Gamblers' Temporary Self-Exclusion from Internet Gambling Sites. *Journal of Gambling Studies*, 1-15. doi.org/10.1007/s10899-018-9782-y
- Challet-Bouju, G., Perrot, B., Romo, L., Valleur, M., Magalon, D., Fatséas, M., ... Hardouin, J-B. (2016). Harmonizing Screening for Gambling Problems in Epidemiological Surveys - Development of the Rapid Screener for Problem Gambling (RSPG). *J Behav Addict, 5*, 239-250. doi.org/10.1556/2006.5.2016.030
- Grall-Bronnec, M., Caillon, J., Humeau, E., Perrot, B., Remaud, M., Guilleux, A., ... Bouju, G. (2016). Gambling among European professional athletes. Prevalence and associated factors. *J Addict Dis, 35*, 278-290. doi.org/10.1080/10550887.2016.1177807

Communications orales

- Perrot, B., Hardouin, J.-B., Grall-Bronnec, & Challet-Bouju, G. (2018).Typology of online gamblers' behaviors : a multilevel latent class cluster analysis applied to player account-based gambling data. *5th International Conference on Behavioral Addictions*, Cologne, Germany, 23-25 April 2018.

Communications affichées

- Perrot, B., Hardouin, J.-B., Grall-Bronnec M., & Challet-Bouju, G. (2018).Typologie des comportements de jeu sur Internet et identification des pratiques à risque : analyse en classes latentes multi-niveaux à partir des données d'un opérateur de jeu en ligne. *12ème Conférence Francophone d'Epidémiologie Clinique (EPICLIN)*, Nice, France, 31 mai-1er juin 2018.
- Perrot, B., Hardouin, J.-B., Reimer, O., Bjerg, O., Grall-Bronnec M., & Challet-Bouju, G. Développement d'un outil d'évaluation des problèmes de jeu liés à la pratique du poker : le Problematic Overuse of Poker Screen (POPS). *10ème édition du Congrès International d'Addictologie de l'ALBATROS*, Paris, France, 1er-3 juin 2016
- Perrot, B., Hardouin, J.-B., Grall-Bronnec, M., Costes, J.-M., & Challet-Bouju, G. (2016). Dépistage des comportements excessifs de jeu sur Internet. *10ème Conférence Francophone d'Epidémiologie Clinique (EPICLIN)*, Strasbourg, France, 25-27 mai 2016.

Autres publications

- Perrot, B., Bataille, E., & Hardouin, J.-B. (2018). validscale : A command to validate measurement scales. Stata Journal, 18(1), 29-50.
- Dirou, S., Chambellan, A., Chevallier, P., Germaud, P., Lamirault, G., Gourraud, P., ... Lemarchand, P. (2018). Deconditioning, fatigue and impaired quality of life in long-term survivors after allogeneic hematopoietic stem cell transplantation. Bone Marrow Transplant, 53, 281-290. doi.org/10.1038/s41409-017-0057-5
- Launey, Y., Lasocki, S., Asehnoune, K., Gaudriot, B., Chassier, C., Cinotti, R., ... Seguin, P. (2017). Impact of Low-Dose Hydrocortisone on the Incidence of Atrial Fibrillation in Patients With Septic Shock. J Intensive Care Med, 885066617696847. doi.org/10.1177/0885066617696847
- Karsenti, D., Coron, E., Vanbiervliet, G., Privat, J., Kull, E., Bichard, P., ... Poincloux, L. (2017). Complete endoscopic sphincterotomy with vs. without large-balloon dilation for the removal of large bile duct stones : randomized multicenter study. Endoscopy, 49, 968-976. doi.org/10.1055/s-0043-114411

- Haddara, S., Jacques, J., Lecleire, S., Branche, J., Leblanc, S., Le Baleur, Y., ... Coron, E. (2016). A novel hemostatic powder for upper gastrointestinal bleeding : a multicenter study (the "GRAPHE" registry). Endoscopy, 48, 1084-1095. doi.org/10.1055/s-0042-116148
- Duchalais, E., Meurette, G., Perrot, B., Wyart, V., Kubis, C., & Lehur, P. (2016). Exhausted implanted pulse generator in sacral nerve stimulation for faecal incontinence : What next in daily practice for patients ? Int J Colorectal Dis, 31, 439-444. doi.org/10.1007/s00384-015-2433-1
- Hardouin, J.-B., Blanchin, M., Feddag, M., Le Néel, T., Perrot, B., & Sébille, V. (2015). Power and sample size determination for group comparison of patient-reported outcomes using polytomous Rasch models. Stat Med, 34, 2444-2455. doi.org/10.1002/sim.6478
- Blanchin, M., Guilleux, A., Perrot, B., Bonnaud-Antignac, A., Hardouin, J.-B., & Sébille, V. (2015). Power and sample size determination for the group comparison of patient-reported outcomes using the Rasch model : impact of a misspecification of the parameters. BMC Med Res Methodol, 15, 21. doi.org/10.1186/s12874-015-0011-4

Table des matières

Liste des abréviations **xvii**

1 Introduction **1**

1.1 Les jeux de hasard et d'argent (JHA) . . . 2

1.1.1 Pratique des JHA en France . . . 2

1.1.2 Définitions et réglementation liées aux JHA . . . 2

1.1.3 Profil des joueurs . . . 4

1.1.4 Le marché des JHA en ligne en 2017 . . . 4

1.1.5 Classification des types de jeu . . . 4

1.2 Problèmes liés à la pratique des JHA . . . 6

1.2.1 Diagnostic des troubles liés au jeu . . . 6

1.2.2 Étiologie . . . 7

1.2.3 Prévalence . . . 9

1.2.4 Problèmes de jeu et jeu de prédilection . . . 9

1.2.5 Problèmes de jeu et Internet . . . 11

1.2.6 Prévention . . . 12

1.2.7 Un problème de santé publique . . . 13

1.3 Motivations et objectifs de la thèse . . . 13

2 Évaluation des problèmes de jeu à partir de données rapportées par les joueurs 15

2.1 Questionnaires utilisés pour le repérage des problèmes de jeu . . . 16

2.2 Harmonisation du dépistage des problèmes de jeu . . . 18

2.3 Le cas particulier du poker . . . 19

2.3.1 Caractéristiques structurelles spécifiques au poker . . . 19

2.3.2 Des conséquences négatives exprimées différemment . . . 19

2.3.3 Pertinence des outils d'évaluation des problèmes de jeu . . . 20

2.3.4 Développement d'un outil d'évaluation des problèmes de jeu spécifiques au poker . . . 22

3 Évaluation des problèmes de jeu à partir de données de comportement réel 29
3.1 Données rapportées vs données de comportement réel 30
3.2 De l'analyse de données rapportées par les joueurs vers l'analyse de données de comportements de jeu réels 30
3.3 Études basées sur l'analyse de données d'activité de jeu 32
3.4 Le projet EDEIN 33
3.4.1 Motivations 33
3.4.2 Article : 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 33
3.4.3 Résumé du protocole 42
3.4.4 Difficultés liées à l'avancement du projet 42
3.4.5 Structure des données de la FDJ 43
3.4.6 Choix méthodologiques et ajustements par rapport au *study protocol* . . . 44

4 Méthodes statistiques 47
4.1 Classification basée sur les modèles à classes latentes 48
4.1.1 Les différentes méthodes de classification non-supervisée 48
4.1.2 Classification basée sur les modèles de mélange (ou sur les classes latentes) 49
4.1.3 Estimation des paramètres 55
4.1.4 Sélection de modèle 56
4.1.5 Modèles à classes latentes multi-niveaux 58
4.2 Modèles mixtes à classes latentes 61
4.2.1 Principe 61
4.2.2 Écriture des modèles 61

5 Caractérisation des comportements de jeu en ligne et repérage des joueurs à risque 65
5.1 Typologie des comportements de jeu en ligne 66
5.1.1 Choix d'analyse 66
5.1.2 Manuscrit : Typology of online lotteries and scratch games gamblers' behaviours : a multilevel latent class cluster analysis applied to player account-based gambling data (accepté pour publication dans *International Journal of Methods in Psychiatric Research*) 68
5.1.3 Principaux résultats 70
5.1.4 Discussion 71
5.2 Analyse longitudinale de l'activité de jeu des nouveaux joueurs 73
5.2.1 Méthodologie 74
5.2.2 Stratégie de sélection de modèle 75

5.2.3 Résultats préliminaires 75

6 Discussion et perspectives **81**

6.1 Discussion 82

6.1.1 Résumé et limites des travaux 82

6.1.2 Discussion sur le jeu responsable 83

6.2 Perspectives 84

6.2.1 Poursuite de l'étude EDEIN 84

6.2.2 Méthode de classification 85

6.2.3 Analyse longitudinale des comportements de jeu 85

6.3 Conclusion 86

Références **87**

Annexes **99**

A Indice Canadien du Jeu Excessif (Ferris & Wynne, 2001) **i**

B Article : Harmonizing Screening for Gambling Problems in Epidemiological Surveys – Development of the Rapid Screener for Problem Gambling (RSPG) **iii**

C Représentation des trajectoires estimées moyennes et des trajectoires individuelles observées **v**

Table des figures

1.1 Montants dépensés annuellement par les joueurs français en 2014 (reproduit de Costes, Eroukmanoff, Richard, & Tovar, 2015) 5

2.1 Items et calcul du score du questionnaire Rapid Screener for Problem Gambling (RSPG) (reproduit de Challet-Bouju et al., 2016) 27

4.1 Exemple de représentation graphique de clusters en fonction de la paramétrisation de la matrice de variance-covariance dans un mélange de lois normales 52

5.1 Trajectoires moyennes estimées de l'argent misé (log euros) pour les sept classes définies par le modèle mixte à classes latentes 78

5.2 Trajectoires moyennes estimées et observées pour les sept classes définies par le modèle 79

C.-1 Trajectoires estimées moyennes et trajectoires individuelles observées viii

Liste des tableaux

1.1 Facteurs de risque associés à une pratique de jeu problématique 10

3.1 Illustration de la structure des données d'activité de jeu 45

4.1 Noms de différents types de modèles à variable latente 53

5.1 Propriétés statistiques des modèles mixtes à classes latentes composés de 1 à 8 classes 76

5.2 Probabilités moyennes d'appartenance aux classes en fonction de la classe la plus probable 76

5.3 Estimations des effets fixes du modèle mixte à sept classes latentes 77

Liste des abréviations

AIC Akaike Information Criterion. 56, 57

ARJEL Autorité de régulation des jeux en ligne. 3, 4, 13, 14, 33, 42, 43, 84, 85

AWE Approximate Weight of Evidence. 57

BIC Bayesian Information Criterion. 56–58, 63, 68, 72, 75, 76, 85

BLRT Bootstrap Likelihood-Ratio Test. 57

CAH Classification Ascendante Hiérarchique. 48, 49

CJA comptes joueurs actifs. 4

CNIL Commission nationale de l'informatique et des libertés. 42, 43

DIF *Differential Item Functioning*. 21, 22

DSM Manuel diagnostique et statistique des troubles mentaux. 6, 16, 17, 20

EDEIN Etude de Dépistage des comportements Excessifs de jeu sur INternet. 13, 14, 42, 44, 83, 84

FDJ Française des jeux. 3

GRCS Gambling-Related Cognitions Scale. 21, 22, 24–26

ICJE Indice Canadien du Jeu Excessif. 16, 17, 20, 21, 24–26, 32, 33, 42, 43, 82, 84

JHA jeu de hasard et d'argent. 2–4, 6–9, 11, 12, 14, 16, 19, 30, 31, 71, 73, 82–84, 86

LCA Analyse en Classes Latentes (*Latent Class Analysis*). 52, 53, 55

LCCA Latent Class Cluster Analysis. 54, 58–60, 66

LMR Lo-Mendel-Rubin Likelihood-Ratio Test. 57

NODS National Opinion Research Center DSM Screen for Gambling Problems. 16, 17

ODJ Observatoire des jeux. 2, 4

POPS *Problematic Overuse of Poker Screen*. 24–26, 82, 83

RSPG Rapid Screener for Problem Gambling. 19, 82

RSPG–I Rapid Screener for Problem Gambling–Interview. 19

RSPG–SA Rapid Screener for Problem Gambling–Self-Assessment. 19

SOGS South Oaks Gambling Screen. 16, 17, 20

TRJ taux de retour au joueur. 3, 4

Chapitre 1

Introduction

Sommaire

1.1 Les jeux de hasard et d'argent (JHA) 2
1.1.1 Pratique des JHA en France 2
1.1.2 Définitions et réglementation liées aux JHA 2
1.1.3 Profil des joueurs 4
1.1.4 Le marché des JHA en ligne en 2017 4
1.1.5 Classification des types de jeu 4
1.2 Problèmes liés à la pratique des JHA 6
1.2.1 Diagnostic des troubles liés au jeu 6
Diagnostic clinique 6
Formes subcliniques 7
1.2.2 Étiologie 7
Modèles théoriques 7
Facteurs de risques 8
1.2.3 Prévalence 9
1.2.4 Problèmes de jeu et jeu de prédilection 9
1.2.5 Problèmes de jeu et Internet 11
1.2.6 Prévention 12
1.2.7 Un problème de santé publique 13
1.3 Motivations et objectifs de la thèse 13

1.1 Les jeux de hasard et d'argent (JHA)

1.1.1 Pratique des JHA en France

La pratique des jeux de hasard et d'argent (JHA) constitue une activité qui occupe une place non négligeable dans les pratiques de loisirs des Français. Dans une enquête menée par l'Observatoire des jeux (ODJ) en 2014 (Costes et al., 2015), 74% des Français âgés de 15 à 75 ans ont déclaré avoir joué au moins une fois à un JHA au cours de leur vie et 56% au cours de l'année écoulée. Parmi les joueurs au cours de l'année écoulée, 53% jouaient entre 1 et 15 fois dans l'année, 32% au moins 52 fois (environ une fois par semaine) et 15% jouaient au moins 104 fois dans l'année (plus d'une fois par semaine).
Selon cette même étude, on estime que 4,1% des Français ont joué au moins une fois à un JHA sur Internet au cours de l'année écoulée. Les types de jeu les plus pratiqués sur Internet étaient les jeux de tirage (pratiqués par 53% des personnes interrogées), le poker (26%), les paris sportifs (24%), les paris hippiques (15%), et les jeux de grattage (14%). Les jeux de casinos et autres jeux, illégaux en France sur Internet, étaient pratiqués par 5% des Français.
D'après une enquête menée en 2017 sur la pratique des JHA sur Internet (Costes & Eroukmanoff, 2018), ces chiffres sont en augmentation puisque, trois ans plus tard, les jeux de loterie étaient pratiqués par 70% des Français, suivis par les jeux de grattage (36%), les paris sportifs (31%), le poker (21%) et les paris hippiques (17%). Les jeux non régulés étaient pratiqués par 19% des personnes interrogées.

1.1.2 Définitions et réglementation liées aux JHA

Les JHA désignent une forme de jeu caractérisée par les trois critères suivants (Ladouceur, 2004a) :

- le joueur mise de l'argent ou un objet de valeur ;
- la mise est irréversible ;
- l'issue du jeu repose principalement sur le hasard.

Les JHA peuvent être rassemblés dans des grandes catégories de jeu :

Jeux de tirage Le joueur mise une somme d'argent et sélectionne des numéros dans une grille. Un tirage est ensuite effectué par l'opérateur pour déterminer les grilles gagnantes.

Jeux de grattage Le joueur achète un ticket et gratte une ou plusieurs cases pour savoir s'il a gagné.

Paris hippiques Le joueur mise de l'argent sur l'issue d'une course hippique.

Paris sportifs Le joueur mise de l'argent sur l'issue d'une rencontre sportive.

Poker Le joueur tente de remporter les jetons des adversaires en constituant la meilleure combinaison de cinq cartes ou en les faisant abandonner. Le montant des jetons de départ est

déterminé par la *cave*, qui correspond à la somme d'argent dépensée par le joueur pour commencer la partie.

Machines à sous Le joueur insère de l'argent dans une machine pour lancer une partie dont l'issue est connue rapidement.

Autres jeux de casino Cette catégorie regroupe les jeux de casino autres que le poker et les machines à sous : la roulette, le Craps, le Blackjack, le Baccara, les autres jeux de cartes et les autres jeux de dés.

Les JHA sont proposés par des opérateurs de jeu, qui prélèvent une partie des mises et redistribuent le reste aux joueurs. On appelle taux de retour au joueur (TRJ) la proportion des sommes d'argent gagnées par les joueurs par rapport à la totalité de leurs mises ou dépenses sur une période donnée.

En France, la réglementation est différente entre les JHA hors-ligne et les JHA sur Internet. Les JHA hors-ligne autorisés en France sont ceux proposés par les trois grands opérateurs de jeu : la Française des jeux (FDJ) (pour les jeux de tirage, jeux de grattage et paris sportifs), le PMU (paris hippiques) et les casinos (poker, machines à sous et autres jeux de casino). La FDJ est sous la tutelle du Ministère de l'Action et des Comptes publics, le PMU est sous la double tutelle du Ministère de l'Agriculture et de l'Alimentation et Ministère de l'Action et des Comptes publics, et l'activité des casinos est sous la double tutelle du Ministère de l'Intérieur et du Ministère de l'Action et des Comptes publics. Depuis la loi n^{o} 2010-476 du 12 mai 2010 relative à l'ouverture à la concurrence et à la régulation des JHA en ligne, les Français peuvent pratiquer légalement trois types de JHA sur Internet : les paris hippiques, les paris sportifs et les jeux de cercle (poker). Pour avoir le droit de proposer un de ces types de jeu, les opérateurs doivent faire une demande d'agrément auprès de l'Autorité de régulation des jeux en ligne (ARJEL) et respecter un cahier des charges spécifique, chaque type de jeu nécessitant un agrément distinct. Créée par la loi du 12 mai 2010, l'ARJEL est une autorité administrative indépendante qui a pour missions de :

- délivrer des agréments et s'assurer du respect des obligations par les opérateurs ;
- protéger les populations vulnérables, lutter contre l'addiction ;
- s'assurer de la sécurité et de la sincérité des opérations de jeu ;
- lutter contre les sites illégaux ;
- lutter contre la fraude et le blanchiment d'argent.

Grâce à une dérogation, la FDJ conserve son monopole historique sur les jeux de tirage et de grattage, et est donc le seul opérateur à proposer ces types de jeu.
En résumé, concernant la règlementation des JHA en ligne, les consommateurs français ont le droit de pratiquer des paris hippiques, des paris sportifs et des parties de poker sur les sites des opérateurs agréés par l'ARJEL, ainsi que les jeux de tirage et de grattage sur le site de la FDJ.

1.1.3 Profil des joueurs

L'enquête de l'ODJ menée en 2014 montre que, comparativement aux non-joueurs, les joueurs de JHA sont plutôt des hommes, âgés de 25 à 54 ans et professionnellement actifs. Les femmes sont davantage représentées parmi les personnes jouant à des jeux de grattage (55%), mais beaucoup moins dans les autres types de jeu (jeux de casino : 29%, paris hippiques : 26%, poker : 19%, paris sportifs : 7%). Comparativement aux personnes ne jouant pas à des JHA sur Internet, les joueurs en ligne sont plutôt des hommes (hommes : 76% parmi les joueurs en ligne *vs* 49% parmi les personnes qui ne jouent pas sur Internet), plus jeunes (moins de 35 ans : 45% *vs* 31%) et sont plus diplômés (diplôme supérieur au bac : 37% *vs* 23%).
La figure 1.1 représente les montants d'argent dépensés par année par joueur pour chaque type de jeu. On note que, tous types de jeu confondus, la moitié des joueurs misent plus de 96 euros par an. Les joueurs de poker sont les plus dépensiers, avec un montant médian de 200 euros.

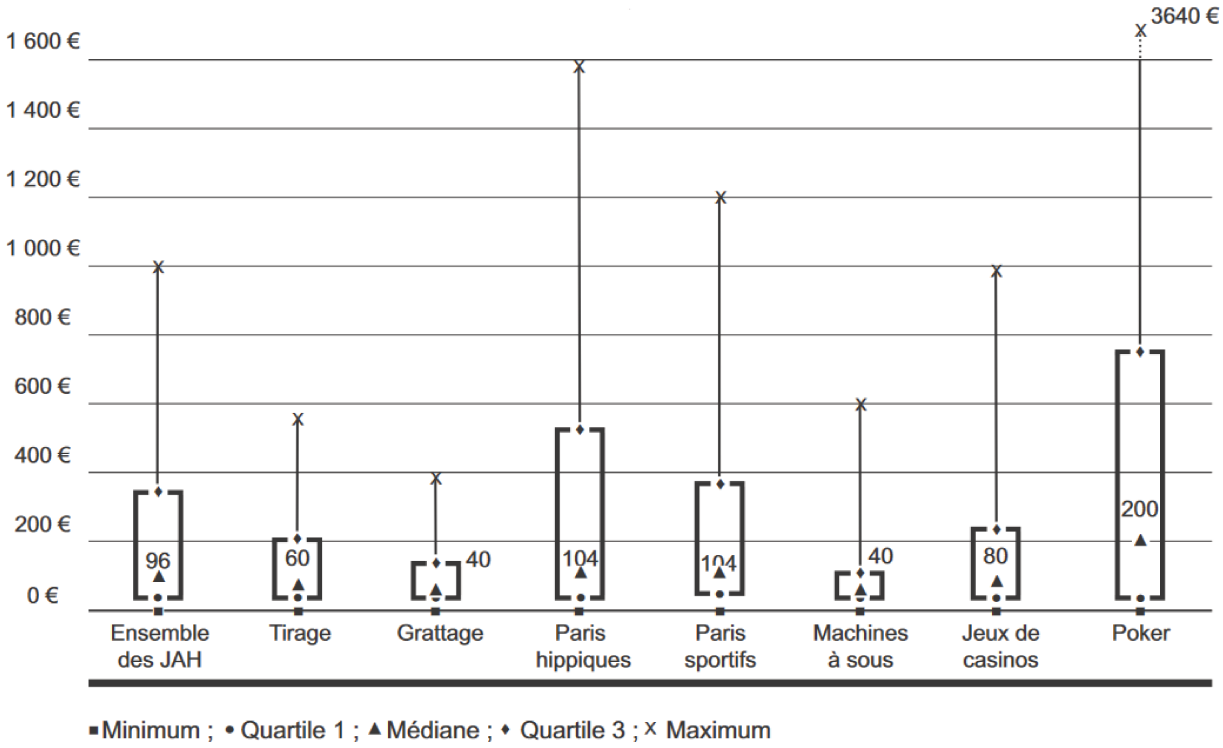


FIGURE 1.1 – Montants dépensés annuellement par les joueurs français en 2014 (reproduit de Costes et al., 2015)

1.1.4 Le marché des JHA en ligne en 2017

L'analyse du marché des jeux en ligne en France pour l'année 2017 effectuée par l'ARJEL montre une augmentation globale de l'activité des JHA en ligne (hors FDJ) au niveau des mises effectuées, du produit brut des jeux (c'est à dire le montant cumulé des mises duquel est déduit

l'argent reversé aux joueurs) et du nombre de comptes joueurs actifs (CJA) (c'est à dire les comptes pour lesquels au moins une prise de jeu a été engagée par le joueur durant la période considérée).
Concernant les paris sportifs, environ 1 993 000 joueurs ont effectué au moins un pari sportif en 2017, pour un total de 2 510 millions d'euros misés. Le TRJ était de 83%.
Concernant les paris hippiques, 523 000 personnes ont pratiqué cette activité, pour un total de 999 millions d'euros et un TRJ de 78%.
Concernant le poker, 1 049 000 joueurs ont participé à au moins une partie, pour un total de mises et de droits d'entrée de 5 642 millions d'euros. Le TRJ était d'environ 96%.
Selon un communiqué de presse de la FDJ, 1 800 000 personnes ont joué en ligne à des jeux de la FDJ en 2017, pour un montant total de 1 700 millions d'euros (y compris les ventes issues des services digitaux). Le TRJ moyen, jeux hors-ligne et en ligne confondus, était de 67%.

1.1.5 Classification des types de jeu

Au delà du regroupement par catégorie de jeu décrit dans le paragraphe 1.1.2, d'autres classifications basées sur les caractéristiques structurelles des JHA ont été proposées dans la littérature. L'une des plus populaires est celle de Boutin (2010) qui a proposé un regroupement en trois classes en fonction de la part d'adresse impliquée dans le jeu et de l'espérance de gain. Cette classification permet de regrouper les types de jeux présentés précédemment de la manière suivante :

1ère classe : Jeux de hasard sans adresse (jeux de hasard pur) : jeux de tirages[1], jeux de grattage et jeux de casino (hors poker et Blackjack)

2ème classe : Jeux de hasard avec quasi-adresse : paris sportifs, paris hippiques, Blackjack

3ème classe : Jeux de hasard avec adresse : poker

Boutin distingue également les jeux qui se jouent "contre la maison" (classes 1 et 2) des jeux qui se jouent contre d'autres joueurs (classe 3). Dans le premier cas, la structure des jeux (par exemple, le nombre de tickets gagnants pour les jeux de grattage ou le système de cotes pour les paris sportifs) rend impossible une espérance de gains positive pour le joueur sur le long terme. Dans le cas où plusieurs joueurs s'affrontent, comme au poker, la faillibilité des joueurs peut permettre à certains d'avoir une espérance de gain supérieure à celle des autres joueurs, donc de gagner de l'argent.

1. Parmi les jeux de tirages, on peut distinguer les *loteries différées*, pour lesquelles le tirage est effectué plusieurs heures ou jours après la validation du ticket, et les *loteries instantanées*, pour lesquelles le tirage intervient quasi-immédiatement.

1.2 Problèmes liés à la pratique des JHA

La majorité des personnes jouant à des JHA le font de manière contrôlée et récréative mais certaines personnes perdent le contrôle de leur pratique de jeu. Le jeu engendre dans ce cas des conséquences négatives d'ordre social, personnel ou professionnel. On parle alors de troubles liés à la pratiques des JHA.

1.2.1 Diagnostic des troubles liés au jeu

Diagnostic clinique

Dans la 5$^{\text{ème}}$ édition du Manuel diagnostique et statistique des troubles mentaux (DSM), les troubles liés aux JHA sont classés dans la section des "troubles liés à l'usage d'une substance et troubles addictifs". Ils partagent ainsi avec les autres addictions deux caractéristiques clés : la perte de contrôle sur le comportement et la poursuite du comportement malgré les conséquences négatives (A. Goodman, 1990). Les critères du DSM-5 définissant les troubles liés aux JHA sont décrits ci-dessous :

A. Pratique inadaptée, persistante et répétée du jeu d'argent conduisant à une altération du fonctionnement ou une souffrance cliniquement significative, comme en témoigne, chez le sujet, la présence d'au moins quatre des manifestations suivantes au cours d'une période de 12 mois :

- 1. Besoin de jouer avec des sommes d'argent croissantes pour atteindre l'état d'excitation désiré
- 2. Agitation ou irritabilité lors des tentatives de réduction ou d'arrêt de la pratique du jeu.
- 3. Efforts répétés mais infructueux pour contrôler, réduire ou arrêter la pratique du jeu.
- 4. Préoccupation par le jeu (ex. préoccupation par la remémoration d'expériences de jeu passées ou par la prévision de tentatives prochaines ou par les moyens de se procurer de l'argent pour jouer).
- 5. Jeu lors de sentiments de souffrance/mal-être (ex. sentiments d'impuissance, de culpabilité, d'anxiété, de dépression).
- 6. Après avoir perdu de l'argent au jeu, retour fréquent au jeu un autre jour pour recouvrer ses pertes (pour se "refaire" (*chasing*)).
- 7. Mensonge en vue de dissimuler l'ampleur réelle de ses habitudes de jeu.
- 8. Mise en danger ou perte d'une relation affective importante, d'un emploi ou de possibilités d'étude ou de carrière à cause du jeu.

- 9. Recours à d'autres personnes pour obtenir de l'argent et se sortir de situations financières désespérées dues au jeu.

B. La pratique du jeu d'argent n'est pas mieux expliquée par un épisode maniaque.

Le trouble est considéré comme léger lorsque 4 à 5 critères sont satisfaits, moyen entre 6 et 7 critères, et grave entre 8 et 9 critères.
Dans un contexte épidémiologique, on préfère cependant s'appuyer sur des questionnaires, remplis de manière autonome par les joueurs, pour dépister les problèmes de jeu et évaluer leur sévérité. Ce point fait l'objet du chapitre 2 du manuscrit.

Formes subcliniques

Outre la classification clinique du trouble, des formes subcliniques sont considérées pour désigner les individus rencontrant des problèmes liés aux JHA, mais pas suffisamment pour poser un diagnostic de trouble addictif avéré. Sans réel consensus, on parle dans ce cas de "joueurs problématiques". Il est à noter que ce terme est parfois utilisé pour désigner de manière globale les individus subissant les conséquences négatives d'un problème de jeu (incluant donc à la fois des formes cliniques et subcliniques). D'autres termes tels que "joueurs excessifs" ou "joueurs à risque" sont également employés dans la littérature pour désigner des individus ayant des problèmes de jeu plus ou moins sévères.

1.2.2 Étiologie

Modèles théoriques

Plusieurs modèles théoriques ont été proposés pour essayer d'expliquer le développement des troubles liés à la pratique des JHA. Ainsi, le modèle le plus consensuel est celui proposé par Blaszczynski et Nower (2002), qui décrivent trois types de *pathways* ("chemins"), correspondant chacun à un profil de joueur problématique :

Les joueurs problématiques au comportement conditionné Il s'agit de joueurs sans psychopathologie prémorbide, dont les troubles liés au jeu se développent autour de facteurs environnementaux et en particulier une augmentation de la disponibilité des JHA. Le joueur développe ensuite des distorsions cognitives liées au jeu qui contribuent au développement puis au maintien de la pratique problématique de jeu, et continue à jouer pour tenter de se refaire.

Les joueurs problématiques émotionnellement vulnérables Également caractérisés par un comportement conditionné, ces joueurs utilisent le jeu comme une échappatoire pour moduler des états affectifs négatifs (stress, anxiété, dépression, etc.) pré-existants, associés à des vul-

nérabilités biologiques (en particulier, dysfonctionnement des systèmes sérotoninergique, noradrénergique et dopaminergique).

Les joueurs problématiques antisociaux-impulsifs Ils partagent les mêmes vulnérabilités biologiques que les individus du groupe 2, mais se distinguent par des traits d'impulsivité et un comportement antisocial se manifestant par des comportements (pas uniquement liés au jeu) inadaptés multiples et sévères.

Cette tentative de typologie théorique des joueurs problématiques se retrouve également dans le modèle de Iancu, Lowengrub, Dembinsky, Kotler, et Dannon (2008) qui suppose trois sous-types de joueurs pathologiques, basés sur la présence de différentes comorbidités psychiatriques :

- le sous-type "addictif", qui représente, selon les auteurs, 50 à 60% des joueurs pathologiques, et qui est plutôt caractérisé par des troubles liés au jeu modérés, une consommation abusive d'alcool et qui ont tendance à prendre de mauvaises décisions. On pourrait rapprocher ce groupe du *pathway* 1 décrit par Blaszczynski et Nower (2002) ;
- le sous-type "obsessif-compulsif" (20-25%), composé de joueurs qui jouent en réponse à un ressenti de traumatisme psychologique (ex : divorce) et caractérisés par des troubles de l'anxiété ou de l'humeur. On pourrait rapprocher ce groupe du *pathway* 2 ;
- le sous-type "impulsif" (10-15%), caractérisé par la présence de comorbidités psychiatriques (troubles du déficit de l'attention et hyperactivité, consommation abusive de substances, etc.), et qu'on pourrait rapprocher des joueurs du *pathway* 3.

De manière similaire, le modèle biopsychosocial de Sharpe (2002) fait une distinction entre, d'une part, les joueurs problématiques caractérisés par une intolérance à l'ennui et des niveaux élevés de recherche de sensation et, d'autre part, les joueurs problématiques qui utilisent le jeu pour échapper à des problèmes ou des événements stressants de leur vie.

Facteurs de risques

D'après le "triangle addictif" d'Olievenstein (1983) proposé dans le contexte de la toxicomanie, l'addiction est conceptualisée comme le résultat d'une rencontre entre un produit, un moment socioculturel et une personnalité. Le modèle a été étendu aux addictions sans substance (Valleur & Matysiak, 2006) et peut s'appliquer dans le contexte de l'addiction aux JHA. Cette approche multifactorielle permet de définir des facteurs de risque à trois niveaux : les facteurs de risque liés au joueur, à l'environnement et aux jeux. L'identification de ces facteurs de risque a fait et fait encore l'objet de nombreuses études. Plusieurs revues de la littérature (ex : Inserm, 2008 ; Johansson, Grant, Kim, Odlaug, & Götestam, 2009 ; Miller, 2015 ; Dowling et al., 2017) ont été réalisées dans ce but. Il s'agit d'un travail complexe en partie à cause de l'hétérogénéité méthodologique des études incluses dans les revues de la littérature. En particulier :

- Les joueurs ne sont pas tous identifiés comme "problématiques" ou "pathologiques" à partir des mêmes outils et des mêmes seuils.

- Les populations sont parfois différentes en termes d'âge ou de genre.
- Différents modes de recrutement sont utilisés : recrutement en population générale, dans des lieux de jeu, dans des structures de soins, etc.
- Les types de jeu pratiqués peuvent différer selon les études et les pays.

Tout en considérant ces limites, le tableau 1.1 résume les facteurs de risque identifiés dans les revues de la littérature précédemment citées.

1.2.3 Prévalence

D'après une revue de la littérature incluant des enquêtes de prévalence réalisées entre 2000 et 2015 (Calado & Griffiths, 2016), la prévalence du jeu problématique dans l'année variait de 2% à 5% en Amérique du Nord, de 0,5% à 5,8% en Asie, de 0,4% à 0,7% en Océanie et de 0,1% à 3,4% en Europe.
En France, d'après l'enquête de prévalence la plus récente (Costes et al., 2015), on estime que 0,5% des Français âgés de 15 à 75 ans sont des "joueurs excessifs" et 2,2% sont des "joueurs à risque modéré". Dans la population des joueurs (i.e. les personnes ayant joué au moins une fois dans l'année), ces pourcentages sont de 0,9% et 3,9%, respectivement.
Dans la population des joueurs en ligne Français, on estime, d'après une enquête menée en 2017, que 13,0% sont des joueurs excessifs et 9,4% sont des joueurs à risque modéré (Costes & Eroukmanoff, 2018). A titre de comparaison, ces pourcentages étaient de 6,6% et 10,4% en 2012 (Tovar, Costes, & Eroukmanoff, 2013).

1.2.4 Problèmes de jeu et jeu de prédilection

Le jeu de prédilection des joueurs serait en partie lié à certains traits de personnalité, notamment la recherche de sensations, la recherche de contrôle sur le jeu et l'atmosphère du jeu (Challet-Bouju et al., 2015). Par exemple, une personne à la recherche de sensations s'orientera davantage vers des jeux comme le poker ou les courses hippiques, un individu recherchant une atmosphère particulière préfèrera aller jouer dans un casino, une personne préférant jouer en solitaire sans volonté de contrôle sur le jeu s'orientera vers des jeux à gratter, etc.
Par ailleurs, la proportion de joueurs problématiques diffère selon le type de jeu le plus pratiqué par les joueurs. Ainsi, le risque de développer des problèmes de jeu semble plus élevé chez les joueurs de machines à sous et les jeux de casino. Les loteries instantanées et le poker apparaissent également associés à des risques de problèmes de jeu plus importants. A l'inverse, les joueurs de loteries différées et de jeux à gratter (hors jeux de grattage sur Internet) semblent moins à risque de développer des problèmes de jeu (Young, Stevens, & Morris, 2008 ; Costes et al., 2011 ; Binde, 2011 ; Binde, Romild, & Volberg, 2017 ; Gainsbury et al., 2014).
Concernant plus spécifiquement les JHA sur Internet, l'étude de Costes et Eroukmanoff (2018) montre également que la proportion de joueurs excessifs varie significativement selon le type de

TABLE 1.1 – Facteurs de risque associés à une pratique de jeu problématique

Facteurs liés à l'individu
Genre masculin
Jeune (adolescents et jeunes adultes)
Délinquance et actes illégaux
Initiation précoce au jeu
Maltraitance dans l'enfance, abus, négligences
Consommation de substances psychoactives
Troubles anxieux et dépressifs
Troubles du déficit de l'attention/hyperactivité
Trouble obsessionnel compulsif
Impulsivité
Niveau élevé de recherche de sensations
Faible estime de soi
Personnalité antisociale
Faible tolérance à l'ennui
Stratégies de coping inadaptées
Dysfonctionnements des systèmes dopaminergiques, noradrénergiques et sérotoninergiques
Facteurs liés à l'environnement et au contexte
Historique de problèmes de jeu dans la famille, apprentissage social
Bas niveau de support social, d'éducation et de revenus
Accessibilité des jeux
Stimulations sensorielles (son, lumière)
Facteurs liés au jeu
Distorsions cognitives liées au jeu (illusion de contrôle, biais d'interprétation, quasi-gains)
Structure du jeu (faible délai mise-gain, forte fréquence de répétition du jeu)
Big Win (Gros gain initial)
Jeu sur Internet
Bonus monétaires versés au joueur

jeu principalement pratiqué. En effet, cette proportion était de 3,6% parmi les joueurs jouant principalement à des loteries à tirage différé, 9,7% pour les jeux de grattage, 10,9% pour les paris hippiques, 10,9% pour les paris sportifs, et 15,6% pour le poker. La proportion de joueurs excessifs était encore plus élevée parmi les personnes jouant sur des sites illégaux : 19,7% pour les jeux de loterie autres que ceux de la FDJ, 45,5% pour les jeux de casino, et 48,3% pour les machines à sous.

1.2.5 Problèmes de jeu et Internet

La pratique des JHA sur Internet présente des particularités pouvant contribuer au développement de comportements de jeu excessifs (Griffiths, 2003 ; Gainsbury, 2015) :

- une plus grande accessibilité aux jeux ;
- l'anonymat ;
- davantage d'options pour les paris ;
- des mises plus importantes ;
- des issues de jeu plus rapides et plus fréquentes ;
- plus de confort ;
- l'utilisation d'une monnaie digitale qui donne l'impression de ne pas dépenser de l'argent réel ;
- des facilités de paiement.

Plusieurs études (ex : Wood & Williams, 2009 ; Griffiths, Wardle, Orford, Sproston, & Erens, 2011 ; Kairouz, Paradis, & Nadeau, 2012 ; Effertz, Bischof, Rumpf, Meyer, & John, 2018 ; Papineau et al., 2018) ont montré que les joueurs en ligne avaient plus de risque d'avoir des problèmes de jeu et/ou des problèmes de jeu plus sévères que les joueurs hors-ligne. D'autres travaux suggèrent que les joueurs jouant à la fois sur Internet et hors-ligne seraient plus à risque que les joueurs jouant exclusivement hors-ligne et également plus à risque que les joueurs jouant exclusivement sur Internet (Gainsbury, Russell, Blaszczynski, & Hing, 2015 ; Papineau et al., 2018).

Par ailleurs, il a été montré qu'après ajustement sur des facteurs liés à la probabilité de jouer sur Internet et/ou à la probabilité d'avoir des problèmes de jeu, le jeu en ligne n'était pas associé à une plus grande sévérité des problèmes de jeu (Wardle et al., 2007 ; Philander & MacKay, 2014 ; Baggio et al., 2017). Cependant, dans ces trois études, le choix des variables d'ajustement était discutable, puisqu'était notamment considéré comme facteur de confusion, l'implication dans jeu (défini comme le nombre de formats de jeu différents ou le temps passé à jouer). Or, comme discuté dans l'article de Baggio et al. (2017), l'implication dans le jeu pourrait constituer un médiateur (et non plus un facteur de confusion) dans la relation entre le jeu en ligne et les problèmes de jeu, Internet rendant les jeux plus accessibles et le nombre de

jeux différents pratiqués étant plus important chez les joueurs en ligne (McCormack, Shorter, & Griffiths, 2013). Il est d'autant plus important de bien déterminer le rôle de cette variable puisque le nombre de formats de jeu pratiqués apparait comme une différence importante lorsqu'on compare des joueurs hors-ligne à des joueurs en ligne (Wood & Williams, 2009).
La nature exacte du lien entre la pratique des JHA sur Internet et l'augmentation supposée du risque de problèmes liés au jeu reste donc encore à définir.

1.2.6 Prévention

On peut distinguer trois types de prévention des problèmes de jeu, selon le public visé (Inserm, 2008) :

La prévention universelle : elle s'adresse à la population générale, sans profil de risque particulier. Il s'agit principalement de messages informant sur les risques possibles liés aux JHA, diffusés à la télévision, à la radio ou en partenariat avec des opérateurs de jeu.

La prévention sélective : elle s'adresse à des sous-populations présentant des facteurs de risque. Cela concerne donc, de manière générale, la population des joueurs ainsi que, de manière plus spécifique, des groupes de joueurs à risque.

La prévention ciblée : elle s'adresse aux personnes présentant un problème de jeu excessif. Il s'agit ici du repérage des joueurs problématiques et d'orientation vers des structures de soins.

Dans le cas des JHA en ligne, les mesures de prévention peuvent prendre la forme de modérateurs de jeu permettant aux joueurs de réguler leur activité de jeu. Cependant, l'efficacité des modérateurs de jeu sur Internet n'a été que très peu étudiée (Caillon, Grall-Bronnec, Challet-Bouju, & Romo, 2017 ; Ladouceur, Shaffer, Blaszczynski, & Shaffer, 2017). Caillon, Grall-Bronnec, Hardouin, Venisse, et Challet-Bouju (2015) ont alors proposé d'évaluer, dans une étude randomisée contrôlée, l'efficacité de quatre modérateurs de jeu : l'auto-exclusion temporaire, l'auto-limitation des mises, la limitation des bonus (bonus de bienvenue et bonus de fidélité), ainsi que l'information sur les risques liés au jeu par l'intermédiaire de *pop-ups*. Les premiers résultats qualitatifs montrent que les modérateurs de jeu sont assez méconnus ou peu utilisés par les joueurs. Concernant l'auto-exclusion d'une durée minimale légale de sept jours, malgré des bénéfices observées à moyen terme, les joueurs ont souligné le manque d'accompagnement dans la démarche d'auto-exclusion (Caillon et al., 2018). Or, d'une manière générale, les programmes de jeu responsable sont d'autant plus efficaces qu'ils sont définis et appliqués entre les joueurs, l'industrie du jeu et les décideurs politiques (Blaszczynski, Ladouceur, & Shaffer, 2004). L'implémentation et l'utilisation des modérateurs des jeu sont donc encore à améliorer.

1.2.7 Un problème de santé publique

Dans un éditorial de 2017 intitulé (“Problem gambling is a public health concern”, 2017), la revue scientifique *The Lancet* notait le manque de preuves solides autour des facteurs de risque associés au jeu problématique pour guider les politiques de santé et la prise en charge des patients. De plus, tout en rappelant l'hétérogénéité de la pathologie, la présence de comorbidités, son caractère épisodique et le manque de connaissance sur la neurobiologie sous-jacente, l'éditorial affirme qu'une "compréhension incomplète n'est pas une excuse pour l'inaction relative au jeu problématique".

En 2018, un éditorial de la revue *Nature* (“Science has a gambling problem”, 2018) faisait les mêmes constats. Le journal déclarait en effet que "le manque de données disponibles pour informer les décideurs politiques et la profession médicale est choquant" et "qu'indépendamment d'un possible manque de délimitations des responsabilités, chaque gouvernement doit agir et assigner une agence ou un ministère approprié au problème du jeu pathologique".

1.3 Motivations et objectifs de la thèse

En 2013, l'ARJEL a rédigé un rapport développant un ensemble de propositions opérationnelles pour lutter contre le jeu excessif et pathologique (ARJEL, 2013). Trente-trois propositions ont ainsi été faites pour traiter les points suivants :

- Amélioration des dispositifs d'information et de sensibilisation des joueurs et du public ;
- Renforcement et adaptation des mécanismes de régulation ;
- Détection, accompagnement et traitement des joueurs problématiques ;
- Mesure de l'importance et des évolutions de la pathologie en France et de l'efficacité des dispositifs de prévention.

Sur le sujet de la détection des joueurs problématiques, l'une des propositions consistait à "Imposer aux opérateurs de mettre en place un dispositif de repérage des joueurs problématiques et pathologiques sur la base d'indicateurs encadrés par l'ARJEL". C'est dans ce cadre qu'a été défini le projet Etude de Dépistage des comportements Excessifs de jeu sur INternet (EDEIN) (NCT02415296), dont l'objectif principal est de développer un modèle de repérage des comportements de jeu en ligne excessifs, en se basant sur des indicateurs recueillis en routine par les opérateurs (argent misé, fréquence de jeu, etc.). Le modèle pourrait ensuite servir de base pour la mise en place du dispositif recommandé par l'ARJEL et ainsi permettre de développer des mesures de prévention adaptées aux profils variés des joueurs.
Les travaux effectués pendant la thèse et présentés dans ce manuscrit s'articulent autour de cet objectif, et portent plus généralement sur les méthodes de repérage et de caractérisation des joueurs excessifs et problématiques.

Dans le chapitre 2, nous nous intéressons aux méthodes d'évaluation des problèmes de jeu (en ligne et hors-ligne) à partir de données rapportées par les joueurs sous forme de questionnaires. Nous traitons d'abord du problème de l'hétérogénéité des instruments de mesures puis de leur adéquation à la population particulière des joueurs de poker.

Le chapitre 3 se concentre sur les JHA en ligne et introduit le changement de paradigme qui consiste à évaluer les problèmes liés aux JHA non plus à partir de données subjectives mais en se basant sur les données de jeu "réelles" liées au compte du joueur. Le protocole de l'étude EDEIN est présenté, ainsi que les ajustements effectués suite aux difficultés techniques et administratives rencontrées au cours du projet.

Le chapitre 4 détaille les méthodes statistiques utilisées pour répondre aux problématiques posées. Il présente dans un premier temps les modèles de classification basés sur les classes latentes et dans un second temps les modèles longitudinaux à variables latentes.

Dans le chapitre 5, nous présentons l'analyse en classes latentes multi-niveaux appliquée aux données d'activité de jeu des joueurs de la FDJ, puis les résultats préliminaires de l'analyse de l'évolution des comportements de jeu durant la période suivant l'ouverture d'un compte à la FDJ.

Les méthodes et les résultats sont discutés dans le chapitre 6. Nous discutons également de la place des travaux effectués et de leurs applications possibles dans le cadre de la prévention des problèmes liés aux JHA.

Chapitre 2

Évaluation des problèmes de jeu à partir de données rapportées par les joueurs

Sommaire

2.1 Questionnaires utilisés pour le repérage des problèmes de jeu . . . 16
2.2 Harmonisation du dépistage des problèmes de jeu 18
2.3 Le cas particulier du poker 19
2.3.1 Caractéristiques structurelles spécifiques au poker 19
2.3.2 Des conséquences négatives exprimées différemment 19
2.3.3 Pertinence des outils d'évaluation des problèmes de jeu 20
2.3.4 Développement d'un outil d'évaluation des problèmes de jeu spécifiques au poker 22
Article en préparation : The Problematic Overuse of Poker Screen (POPS) questionnaire : a new tool to explore gambling problems in poker 22
Résumé de la méthodologie 24
Résultats principaux 24
Discussion 25

2.1 Questionnaires utilisés pour le repérage des problèmes de jeu

L'évaluation clinique à l'aide du DSM-5 constitue la référence pour le diagnostic des troubles liés à la pratique des JHA. Cependant, en dehors du cadre clinique, d'autres formes d'évaluations ont été proposées. Il s'agit de questionnaires composés de plusieurs questions (items) auxquelles les joueurs répondent parmi les modalités de réponse proposées. Des poids sont affectés aux modalités de réponse (ex : jamais=0, parfois=1, souvent=2) puis un score total est calculé. Des seuils sont souvent définis à partir du score pour définir des groupes de joueurs (ex : joueur sans risque de problèmes de jeu, joueur à risque modéré, joueur à risque élevé, etc.).
On distingue les hétéro-questionnaires qui sont remplis par un évaluateur suite aux réponses données par les joueurs et les auto-questionnaires qui sont remplis par les joueurs eux-mêmes.
L'un des hétéro-questionnaires les plus utilisés est le National Opinion Research Center DSM Screen for Gambling Problems (NODS) (Gerstein et al., 1999 ; Hodgins, 2004). Le NODS est composé de 17 items permettant d'évaluer les 10 critères du DSM-4 (la différence principale entre les critères du DSM-4 et du DSM-5 est l'abandon du critère "actes illégaux" dans la version la plus récente). Il a été développé pour être utilisé en tant qu'outil de dépistage des problèmes de jeu à partir d'entretiens téléphoniques en population générale.

Par rapport aux hétéro-questionnaires, les auto-questionnaires ont l'avantage de nécessiter des ressources moins coûteuses (notamment en temps et en argent). Les auto-questionnaires plus fréquemment utilisés sont décrits ci-dessous.

Le South Oaks Gambling Screen (SOGS) (Lesieur & Blume, 1987) est composé d'une première partie permettant de connaitre les fréquences de pratique de différents types de jeu. Dans la seconde partie, deux questions portent sur les montants d'argent joués et les antécédents familiaux liés au jeu. Le score est calculé à partir des 20 questions suivantes qui portent sur le *chasing* (le fait de rejouer pour regagner de l'argent perdu), les répercussions négatives du jeu, le ressenti d'une pratique problématique, les tentatives d'arrêt et les moyens de financement. Un score de 0 indique l'absence de problème de jeu, un score compris entre 1 et 4 indique la présence de quelques problèmes de jeu, et un score supérieur à 5 permet d'identifier un joueur pathologique probable.

Le NODS-SA est la version "auto-évaluation" du NODS. Cette version n'a jamais fait l'objet d'une validation.

Le NODS–CLiP (Toce-Gerstein, Gerstein, & Volberg, 2009) est composé de 3 items sélectionnés parmi les 17 qui constituent le NODS. Bien que développé à partir d'enquêtes téléphoniques, ses auteurs affirment que le questionnaire peut être auto-administré. Les 3 items portent sur le contrôle, le mensonge et la préoccupation. Un joueur répondant positivement à au moins une des 3 questions est considéré comme un joueur pathologique.

L'Indice Canadien du Jeu Excessif (ICJE) (Ferris & Wynne, 2001) est composé de 9 items à

4 modalités de réponses et met l'accent sur les conséquences négatives du jeu (ex : "Vos habitudes de jeu ont-elles causé des difficultés financières à vous ou votre entourage ?", "Vous êtes vous déjà senti coupable de vos habitudes de jeu ou de ce qui arrive quand vous jouez ?"). Le questionnaire ICJE est en fourni Annexe A. Le score total permet de classer les joueurs dans les catégories suivantes : joueur non problématique (ICJE=0), joueur à risque faible (ICJE=1–2), joueur à risque modéré (ICJE=3–7) et joueur problématique (ICJE=8–27). Remettant en cause la validité des deux catégories intermédiaires, Currie, Hodgins, et Casey (2013) ont proposé de classer dans le groupe des joueurs à risque faible les répondants ayant un score compris entre 1 et 4, et de classer dans le groupe des joueurs à risque modéré les répondants ayant un score compris entre 5 et 7.

Le Questionnaire Lie/Bet (Johnson et al., 1997) est composé de 2 items portant sur le mensonge à propos de la pratique de jeu et sur le besoin de miser des sommes d'argent de plus en plus importantes. Une réponse positive à au moins une des 2 questions indique un joueur problématique probable.

Ces questionnaires peuvent être classés en trois catégories selon leur nombre d'items (Thomas et al., 2011) : les questionnaires courts (1–3 items), comme le NODS-CLiP et le Questionnaire Lie/Bet, les questionnaires de taille moyenne (4–12 items), comme l'ICJE, et les questionnaires longs (>13 items), comme le SOGS.
De plus, tous les questionnaires n'ont pas la même finalité. Thomas, Jackson, et Blaszczynski (2003) ont identifié cinq objectifs majeurs :

- le "diagnostic" d'une pratique de jeu problématique ;
- l'évaluation de la sévérité des problèmes des jeu et de leur conséquences ;
- la prédiction de problèmes de jeu futurs (identification de joueurs à risque de développer des problèmes de jeu) ;
- l'identification de besoins spécifiques pour traiter les problèmes de jeu ;
- le repérage ou le dépistage de problèmes de jeu pour orienter le joueur vers une évaluation plus approfondie.

Néanmoins, les auteurs du rapport ont noté que ces objectifs sont rarement définis et que les questionnaires sont souvent utilisés de manière interchangeable à des buts différents.

Concernant la validation des outils de repérage, la sensibilité (probabilité de détecter les joueurs ayant effectivement des problèmes de jeu) et la spécificité (probabilité de détecter les joueurs n'ayant pas de problèmes de jeu) sont les critères les plus importants. La mesure de référence permettant de déterminer le statut réel des joueurs (en théorie, l'évaluation par un clinicien à l'aide des critères du DSM) n'est cependant pas constante d'une validation à l'autre. Par exemple, les capacités diagnostiques du NODS–CLiP ont été mesurées en prenant comme référence le score total du NODS.

Dans le cas des outils davantage destinés à quantifier et qualifier les problèmes de jeu et/ou leurs conséquences, l'évaluation des propriétés psychométriques semble pertinente, d'autant plus qu'il s'agit généralement de questionnaires plus longs et donc plus adaptés à ce genre d'analyses. En substance, les propriétés psychométriques à évaluer sont les suivantes :

La validité de contenu mesure à quel point les items reflètent le concept d'intérêt.

La validité de critère consiste à mesurer le lien entre le score à tester et un score existant mesurant le même concept.

La validité de construit s'intéresse à la corrélation entre, d'une part, les différents items, et d'autre part, la corrélation entre les items et le score total. Elle inclut la validité de structure qui consiste à étudier si les items mesurent un seul ou plusieurs concepts et la *known-groups validity*, qui consiste à analyser les différences de scores entre des groupes de répondants définis de sorte qu'on observe en théorie des différences importantes.

La fiabilité teste la variabilité (le "bruit") liés aux mesures. Plus cette variabilité est faible, plus le questionnaire est fiable. La fiabilité inclut les concepts de consistance interne et de reproductibilité.

La reproductibilité teste si les mesures sont stables au cours du temps, pour un état inchangé.

La sensibilité au changement mesure la capacité du questionnaire à détecter des changements significatifs du point de vue du domaine d'intérêt.

Une plus large description des propriétés psychométriques est disponible dans l'ouvrage de Fayers et Machin (2013) et celui de de Vet, Terwee, Mokkink, et Knol (2011). La commande Stata - validscale (Perrot, Bataille, & Hardouin, 2018) permet de réaliser de manière simple les analyses quantitatives requises lors de l'évaluation de ces propriétés.
En pratique, la majorité des questionnaires présentés précédemment étant utilisés à des fins de repérage ou de dépistage, la qualité des propriétés psychométriques des outils développés restent secondaires par rapport à la performance mesurée par la sensibilité et la spécificité (Thomas et al., 2011).

2.2 Harmonisation du dépistage des problèmes de jeu

Devant l'hétérogénéité des outils de dépistage des problèmes de jeu existants et le manque de rigeur dans leur validation (en particulier pour les questionnaires courts) (Stinchfield, McCready, & Turner, 2012), nous avons proposé de développer deux questionnaires (un auto-questionnaire et un hétéro-questionnaire), composés de trois items ou moins, à partir de l'étude des meilleures combinaisons (en termes de sensibilité et spécificité) des items provenant des outils les plus utilisés dans ce contexte (Challet-Bouju et al., 2016). L'article présentant ce travail est en Annexe B.

L'outil final, le Rapid Screener for Problem Gambling (RSPG), est composé d'une version Interview (Rapid Screener for Problem Gambling–Interview (RSPG–I)) à utiliser pour les enquêtes téléphoniques ou en face à face et d'une version Auto-Évaluation (Rapid Screener for Problem Gambling–Self-Assessment (RSPG–SA)) destinée aux enquêtes par Internet. La figure 2.1 décrit les items et le calcul des scores issus des deux versions. Le RSPG–I a une sensibilité de 95%

RSPG-I – version Interview (pour les enquêtes téléphoniques ou en face-à-face)

Avez-vous eu une pratique de jeu au cours des 12 derniers mois?

Oui	Continuer l'interview
Non	Arrêter l'interview – 0 point

Durant les 12 derniers mois, avez-vous eu besoin de jouer avec des sommes d'argent de plus en plus élevées pour atteindre l'état d'excitation désiré?

Oui	1 point
Non	0 point

Durant les 12 derniers mois, avez-vous fait des efforts répétés mais infructueux pour contrôler, réduire ou arrêter la pratique du jeu?

Oui	1 point
Non	0 point

Cotation et interprétation:

Score de 0: pas de besoin d'une interview plus poussée ⇒ pas de trouble lié au jeu

Score supérieur ou égal à 1: faire une interview plus poussée pour établir un diagnostic final de trouble lié au jeu

RSPG-SA – version Auto-Evaluation (pour les enquêtes par Internet)

Avez-vous eu une pratique de jeu au cours des 12 derniers mois?

Oui	Continuer le questionnaire
Non	Arrêter le questionnaire – 0 point

Lorsque vous avez joué au cours des 12 derniers mois, combien de fois êtes-vous retourné au jeu un autre jour pour vous refaire, c'est-à-dire pour regagner l'argent perdu auparavant?

Jamais	0 point
Quelquefois (moins de la moitié des fois où j'ai perdu)	0 point
La plupart des fois où je perds	1 point
A chaque fois que je perds	1 point

Au cours des 12 derniers mois, vous êtes-vous déjà senti coupable à cause de la façon dont vous jouez ou à cause de ce qui se produit lorsque vous jouez?

Oui	1 point
Non	0 point

Au cours des 12 derniers mois, avez-vous envisagé d'arrêter de jouer tout en pensant que vous en étiez incapable?

Oui	1 point
Non	0 point

Cotation et interprétation:

Score de 0: pas de besoin d'une interview plus poussée ⇒ pas de trouble lié au jeu

Score supérieur ou égal à 1: faire une interview plus poussée pour établir un diagnostic final de trouble lié au jeu

FIGURE 2.1 – Items et calcul du score du questionnaire Rapid Screener for Problem Gambling (RSPG) (reproduit de Challet-Bouju et al., 2016)

et une spécificité de 78% ; le RSPG–SA a une sensibilité de 95% et une spécificité de 64%. Le fait d'avoir privilégié la sensibilité pour capturer la grande majorité des joueurs problématiques explique la faible spécificité du RSPG. Ce "dépistage" doit donc être suivi d'une évaluation supplémentaire, afin de ne pas surévaluer la proportion de joueurs problématiques. Néanmoins, un premier repérage avec un questionnaire court tel que le RSPG permet une économie importante de temps et de ressources dans les études de dépistage des problèmes de jeu à grande échelle.

2.3 Le cas particulier du poker

Les outils d'évaluation des problèmes de jeu présentés précédemment ont été conçus et sont utilisés indépendamment des types de jeu pratiqués par les répondants. Cependant, on peut se demander s'ils sont adaptés à toutes les formes de jeu. En particulier, le poker possède des caractéristiques structurelles qu'on ne retrouve pas dans les autres JHA. Ces spécificités pourraient expliquer les liens particuliers qui existent entre les problèmes de jeu et la pratique du poker, par rapport aux autres jeux.

2.3.1 Caractéristiques structurelles spécifiques au poker

Le poker est un jeu caractérisé par une part d'adresse (ou habilité) plus importante que dans les autres JHA. Autrement dit, il est possible d'acquérir de l'expérience et des compétences permettant d'espérer un gain d'argent grâce à l'entraînement. De plus, le poker oppose uniquement les joueurs entre eux, sans l'intervention active d'un opérateur ; tous les joueurs sont donc potentiellement faillibles. Au delà de l'aspect financier du jeu, une motivation à jouer au poker pourrait être la volonté de se confronter à d'autres joueurs et essayer de leur être "supérieur", ce qui est relativement spécifique de ce type de jeu. Par ailleurs, ces caractéristiques propres au poker permettent notamment à certains joueurs de générer des revenus significatifs et d'en faire une activité professionnelle ou semi-professionnelle, ce qui n'est pas possible pour les autres formes de jeu.

2.3.2 Des conséquences négatives exprimées différemment

Contrairement à ce qui est généralement observé pour les autres JHA, les problèmes financiers liés au poker ne semblent pas toujours être une conséquence majeure d'une pratique de jeu problématique (Shead, Hodgins, & Scharf, 2008 ; Bjerg, 2010). A l'inverse, les conséquences psychologiques semblent plus importantes, notamment les troubles de l'humeur (anxiété et dépression), le stress et les troubles du sommeil (Barrault, Untas, & Varescon, 2014 ; Eroukmanoff, Costes, & Tovar, 2014 ; Hopley & Nicki, 2010 ; Moreau, Chabrol, & Chauchard, 2016). Les joueurs de poker manifestent également davantage de problèmes de jeu en termes de temps excessif passé à jouer (Shead et al., 2008 ; Hopley & Nicki, 2010). De plus, alors que pour les autres formes de jeu, les distorsions cognitives influent à la fois sur la survenue et le maintien des problèmes de jeu, il semblerait que, dans le cas du poker, elles jouent un rôle plus important dans le déclenchement des problèmes de jeu que dans leur maintien (Mitrovic & Brown, 2009 ; Wood, Griffiths, & Parke, 2007). Toutefois, la validité de l'évaluation des distorsions cognitives dans le cadre du poker est discutable puisque la plupart des outils évaluant les distorsions cognitives liées au jeu ont été développés en références à des jeux sans ou avec quasi-adresse.

2.3.3 Pertinence des outils d'évaluation des problèmes de jeu

Plusieurs études ont souligné l'importance de prendre en considération les spécificités du poker lors de l'évaluation des problèmes de jeu parmi les joueurs de poker (Bouju, Grall-Bronnec, Quistrebert-Davanne, Hardouin, & Vénisse, 2013 ; Laakasuo, Palomäki, & Salmela, 2015 ; Moreau et al., 2016 ; Wood et al., 2007). En effet, les instruments de mesure habituellement utilisés (ex : l'ICJE, le SOGS ou les questionnaires basés sur les critères du DSM) sont certes utiles de manière générale, mais certains items ne seraient pas ou peu adaptés au cas du poker et certains aspects spécifiques au poker ne seraient pas pris en compte.
Premièrement, comme mentionné précédemment, le temps passé à jouer semble très corrélé à une pratique excessive du poker (Hopley & Nicki, 2010 ; Hopley, Dempsey, & Nicki, 2012). Or, la notion de temps passé à jouer n'est présent ni dans l'ICJE, ni dans le SOGS, qui sont les deux outils les plus utilisés.
Deuxièmement, une pratique problématique du poker semble associée à un faible score de "satisfaction de vie" (Laakasuo et al., 2015). Il a également été suggéré que certains personnes jouent au poker pour expérimenter des sensations qu'elles ne parviendraient pas à ressentir dans leur vie quotidienne (Bouju et al., 2013). Cette dimension émotionnelle forte, que l'on retrouve aussi dans l'étude de Barrault et al. (2014), est cependant absente de la plupart des questionnaires.
Troisièmement, certains items du SOGS et de l'ICJE (ex : "Vos habitudes de jeu ont-elles causé des difficultés financières à vous ou à votre entourage ?") font référence aux problèmes financiers, qui, nous l'avons vu, ne font pas nécessairement partie des problèmes les plus importants chez les joueurs de poker.
Enfin, l'évaluation des distorsions cognitives chez les joueurs de poker représente également un challenge. D'une façon générale (pour l'ensemble des formes de jeu), les distorsions cognitives liées au jeu correspondent à des pensées irrationnelles ou erronées portant principalement sur la compréhension du hasard et le contrôle que le joueur peut avoir sur les résultats du jeu (Ladouceur, 2004b). Les distorsions cognitives liées au jeu jouent un rôle majeur dans le développement et le maintien des problèmes de jeu (Ladouceur, 2004a ; Raylu & Oei, 2004), aussi est-t-il important de les évaluer, en plus des symptômes et des conséquences des problèmes de jeu (mesurés avec l'ICJE par exemple). Les distorsions cognitives liées au jeu sont souvent mesurées à l'aide du questionnaire Gambling-Related Cognitions Scale (GRCS) développé par Raylu et Oei (2004). Le GRCS explore cinq dimensions :

Les espoirs liés au jeu (exemple d'item : "Jouer rend l'avenir plus prometteur") ;

L'illusion de contrôle (exemple d'item : "Prier m'aide à gagner") ;

Le contrôle prédictif (exemple d'item : "Les pertes au jeu doivent être suivies par une série de gains") ;

L'incapacité à arrêter de jouer (exemple d'item : "Mon désir de jouer est tellement plus fort que moi") ;

Les biais d'interprétation (exemple d'item : "Relier mes pertes à de la malchance ou de mauvaises circonstances me fait continuer à jouer").

Or, parmi les 23 items du GRCS, certains pourraient être inadaptés à un jeu faisant intervenir une réelle adresse, tel que le poker. Par exemple, selon Lévesque, Sévigny, Giroux, et Jacques (2017), on peut *a priori* penser que sept items pourraient poser problème : deux items liés aux aptitudes ("Relier mes gains à mon adresse et mes capacités me fait continuer à jouer", "Je possède une certaine capacité à prédire mes gains au jeu"), deux items liés aux probabilités ("Relier mes pertes aux probabilités me fait continuer à jouer", "Relier mes pertes à de la malchance ou de mauvaises circonstances me fait continuer à jouer") et trois items liés à l'expérience ("Une série de pertes me procurera un apprentissage qui m'aidera à gagner par la suite", "Jouer rend l'avenir plus prometteur", "Lorsque je gagne une fois, je gagnerai sûrement encore"). A titre d'illustration, être d'accord avec cette dernière affirmation serait un signe de distorsion cognitive pour un joueur de loterie, les événements de jeu étant indépendants entre eux. Par contre, dans le cas du poker, remporter une main constitue souvent un avantage pour la suite de la partie ; dans ce cas, être d'accord avec la proposition ne devrait donc pas être forcément considéré comme une pensée erronée. Les scores du GRCS seraient donc biaisés pour les joueurs de poker.
Dans le but de valider ces hypothèses et de quantifier les éventuels biais, Lévesque et al. (2017) ont analysé le fonctionnement différentiel des items (*Differential Item Functioning* (DIF)) du GRCS. Le principe du DIF est le suivant : les items d'un questionnaire bien construit doivent en théorie être tous compris et interprétés de la même manière, indépendamment des caractéristiques des répondants ou de groupes similaires de répondants. Si pour un même niveau de distorsions cognitives, deux groupes d'individus ne répondent pas, en moyenne, de la même façon à un item donné, alors on parle de fonctionnement différentiel pour cet item. En reprenant l'exemple de l'item "Lorsque je gagne une fois, je gagnerai sûrement encore", l'hypothèse en cas de DIF est qu'un joueur de hasard pur n'ayant pas ou peu de pensées erronées liées au jeu réponde qu'il est en désaccord avec l'affirmation, alors qu'un joueur de poker présentant également peu ou pas de pensées erronées pourrait déclarer être d'accord. Dans ce cas, l'item ne fonctionnerait pas de la même manière entre le groupe des joueurs de poker et le groupe des joueurs pratiquant un autre type de jeu. Les joueurs de poker auraient donc un score de distorsions cognitives surestimé. Au final, les auteurs de l'étude ont identifié un DIF élevé pour trois items : "Relier mes gains à mon adresse et mes capacités me fait continuer à jouer", "Une série de pertes me procurera un apprentissage qui m'aidera à gagner par la suite" et "Relier mes pertes aux probabilités me fait continuer à jouer". Il semblerait donc bien que les distorsions cognitives des joueurs de poker soient mal évaluées avec le GRCS.

2.3.4 Développement d'un outil d'évaluation des problèmes de jeu spécifiques au poker

Conscients des problèmes évoqués précédemment, des chercheurs ont développé des questionnaires pour étudier les aspects spécifiques du poker (Barrault & Varescon, 2013 ; Wood et al., 2007). Cependant, chacun de ces questionnaires a été développé et utilisé uniquement par leurs auteurs, dans une même étude, sans être validé. De plus, Moreau et al. (2016) ont recommandé la création et l'évaluation d'un outil permettant d'étudier précisément les caractéristiques des joueurs de poker. Nous nous sommes donc proposés de développer un tel outil.
En 2014, Ole Bjerg a développé une version préliminaire d'un questionnaire en se basant sur son expérience et des interviews de joueurs de poker. Suite à un changement de poste, il n'a pas pu poursuivre ce projet et a accepté que nous finissions le travail de validation du questionnaire. L'article en préparation relatif à cette étude est présenté ci-dessous.

Article en préparation : The Problematic Overuse of Poker Screen (POPS) questionnaire : a new tool to explore gambling problems in poker

The Problematic Overuse of Poker Screen (POPS) questionnaire: a new tool to explore gambling problems in poker

Bastien Perrot[*,a,b], Jean-Benoit Hardouin[a], Olivier Reimer[b], Ole Bjerg[c], Marie Grall-Bronnec[a,b], Gaëlle Challet-Bouju[a,b]

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

Address: U1246 SPHERE - IRS2, 22 boulevard Bénoni Goullin – 44200 Nantes, France

[b]CHU Nantes, Clinical Investigation Unit "Behavioral Addictions/Complex Affective Disorders", Department of Addictology and Psychiatry, Nantes, France

Address: IFAC - CHU de Nantes - Hôpital St Jacques - Bâtiment Louis Philippe - 85 rue St Jacques 44093 NANTES cedex 1

[c]Copenhagen Business School, Department of Management, Politics and Philosophy, Copenhagen, Denmark

Address: Department of Management, Politics and Philosophy - Copenhagen Business School - Porcelænshaven 18B, 2000 Frederiksberg, Denmark

[*] bastien.perrot@univ-nantes.fr; +33 253009129

Abstract

Poker is a gambling game that differs from others, in particular as it requires special skill. Although several validated tools assessing gambling disorders in a general manner exist, they are often not adapted for poker. The primary objective of this study was to develop and to validate a new tool to assess gambling problems specifically linked to excessive poker practice: the Problematic Overuse of Poker Screen (POPS) questionnaire. Participants were recruited in French poker clubs. Items were selected from a preliminary 20-item version, based on psychometric properties and clinical relevance. The secondary objective was to describe and compare poker players according to a cut-off defined for the POPS. The questionnaire comprised six items, exhibited moderate psychometric properties and seems to bring a better discrimination of current problem poker players than the Problem Gambling Severity Index. The secondary objective showed that problem poker players played more aggressively (lot of bets and raises) than non-problem poker players and played a larger number of other games. Further validation is needed to better assess criterion validity and to provide a validation of an English version. The development of the POPS provides a new useful tool for identifying problem gamblers among poker players.

Keywords

Gambling disorder, poker, scale development, scale validation

1. Introduction

Gambling is a worldwide leisure activity. The majority of gamblers engage in controlled and recreational gambling practice, but some lose control and develop excessive practice. The Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5), defines gambling disorder as a "persistent and recurrent problematic gambling behavior leading to clinically significant impairment or distress" (American Psychological Association, 2013) associated with negative personal, professional or financial consequences. In 2015, a French national survey estimated that 2.7% of the population could be considered as problem gamblers (Costes et al., 2015).

Gambling games can be classified according to the proportions of chance and skill in the game (Bjerg, 2010; Boutin, 2010). Thus, we can distinguish between (i) pure chance gambling games where the player's skill cannot influence the outcome of the game (e.g. slot machines, lotteries, scratch games, etc.); (ii) gambling games with an element of skill, where the outcome can be influenced by competence or knowledge, but with no positive expected value in the long term. In this case, the gambler plays against a gambling industry programmed to be profitable in the long term (e.g. sports betting, Blackjack); (iii) gambling games where the outcome can be significantly influenced by the player's skill, which can lead to positive expected value in the long term since the gambler is playing against other potentially fallible gamblers (e.g. poker).

Poker differs from other gambling games in particular as it requires skill and experience, but also as it is a game played against other players, thus leading to specific problems compared with other gambling games. For example, Shead et al. (2008) showed that poker players spend more time gambling compared to non-poker gamblers. Similarly, Eroukmanoff et al. (2014) found that 45.6% of a sample of online poker players acknowledged that poker interferes with their sleep time (vs. 13.0% for other online gamblers). Shead et al. (2008) added that problem poker players might not always exhibit severe financial problems, which might make them less likely to admit they have gambling related problems. Another study (Mitrovic & Brown, 2009) showed that distorted cognitions may play a less important role in the maintenance of gambling problems for poker players compared to other forms of gambling. Moreover, Bouju et al. (2013) suggested that problem gamblers were likely to take on another persona (an imaginary character created for poker) while playing poker, allowing them to escape the real world if their personal or professional life is unsatisfactory.

Several studies (Bouju et al., 2013; Laakasuo et al., 2015; Moreau et al. 2016; Wood et al., 2007) highlighted the need to take the specificities of poker in particular into account when assessing gambling problems among poker players. The tools generally used by researchers in such situations include for example the Problem Gambling Severity Index (PGSI) (Ferris & Wynne, 2001), the South Oaks Gambling Screen (SOGS) (Lesieur & Blume, 1987) or DSM-based questionnaires. Despite being very useful, these questionnaires sometimes contain items which are not adapted for poker players or do not include important elements specific to poker overuse. For example, as emphasized by Hopley (Hopley et al., 2012; Hopley & Nicki, 2010), problem gambling was correlated with time spent gambling. Another example is that problem poker gamblers could use poker to experience powerful emotions they cannot experience in other areas of life (Bouju et al., 2013), which is consistent with the fact that a higher problem gambling severity is correlated with a lower satisfaction of life (Laakasuo, 2015). However, most of these questionnaires do not include neither the strong emotional dimension often found among poker players (Barrault et al., 2014), nor the time spent gambling. Moreover, Laakasuo et al. (2015) argued that the PGSI and the SOGS "*do not adequately take into consideration in their items the possibility that games of skill, such as poker, might entail behavioral dynamics dissimilar to those observed in games of chance*". In particular, they found that level of poker experience (assessed by the number of years of practice, level of stakes played and the estimated number of hands played during life), despite positively correlated with PGSI, was not associated with lower levels of well-being (even for those playing for long hours in apparent social isolation) but rather was associated with proficient emotion regulation abilities. Furthermore, it should be noted that three of the nine criteria of the gambling disorder section of the DSM-5 are related to financial matters, whereas it is not necessarily the biggest problem encountered by poker players. The PGSI also includes such items related to financial problems (e.g. "Has your gambling caused any financial problems for you or your household?").

Questionnaires have been compiled by researchers to investigate specific aspects of poker, but they were used in their own studies without being validated (Barrault & Varescon, 2013; Wood et al., 2007). Moreau et al. (2016) recommended developing and validating a specific tool for online poker players in order to provide for the in-depth study of the characteristics of online poker players. They added that "*the properties of psychometric tools, particularly regarding excessive gambling and irrational beliefs, should be adapted to poker in order to answer the questions raised by several studies*". Bjerg had developed a preliminary version of such a tool, based on clinical experience and interviews of poker gamblers, which has never been validated.

In this study, we proposed the development of a new version of this tool for the rapid assessment of gambling problems in the specific case of poker. This new version was named the Problematic Overuse of Poker Screen (POPS) questionnaire. It can be used to harmonize evaluation of the poker player population and to prevent over- or under- estimation of gambling problems among poker players. The secondary objective was to compare the characteristics of poker players considered to be problematic to those considered to be non-problematic according to the POPS.

2. Methods

2.1 Procedure

Data was collected between February 17th and November 27th, 2015, via an anonymous Internet survey. Participants were recruited through French poker clubs or federations of poker clubs and from the database of research volunteers compiled by the Federative Institute of Behavioral Addictions (Nantes, France). All adult gamblers who reported gambling on poker on at least one occasion during the previous year were eligible for the study. Participants were effectively included if they completed at least the poker-specific questionnaire (the POPS, see methods section) and the PGSI.

2.2 Participants

The survey was completed by 1,063 participants. After exclusion of non-eligible participants and those who did not answer both the POPS and the PGSI, 718 participants remained in the final sample. Participants' sociodemographic characteristics are given in Table 1. Participants were mainly employed, male and living as a couple. The age ranged from 18 to 80. According to the PGSI scores, the sample comprised 270 (37.6%) non-problem gamblers, 247 (34.4%) low-risk gamblers, 180 (25.1%) moderate-risk gamblers and 21 (2.9%) problem gamblers.

[Insert Table 1 here]

2.2 Measures

2.2.1 Sociodemographic characteristics A short questionnaire included a few questions about gender, age, educational level, occupation, living conditions and level of income.

2.2.2 Gambling habits Participants were asked about their participation in various forms of gambling over the past year (a derived measure of involvement – i.e. number of different games

played – was then calculated), age at which they were introduced to gambling, if a big win occurred during their first poker experiences and poker gambling habits (frequency, money wagered in poker monthly, length of poker sessions, variants played, strategies used, etc.). They were also asked about their perception of poker as a game of strategy, chance or both and whether poker was their favorite gambling activity. Finally, several questions were asked about potential gambling problems encountered by the participants including the feeling of having a gambling problem and at which age, considering stopping gambling because of this problem, having started treatment for a gambling problem and at which age, abstinence of gambling during at least one month (all gambling types), self-assessment of self-efficacy to control gambling urges ("How would you rate your ability to control your gambling urges?") and self-assessment of damage measured on a 10-point visual scale ("How much does your gambling behavior disrupt these different areas of your life? : family, friends, studies/work, health and finances").

2.2.3 Problem Gambling Severity Index (PGSI) The PGSI is a 9-item self-assessment questionnaire (Cronbach's alpha (α) = 0.84) within the Canadian Problem Gambling Index (Ferris & Wynne, 2001). It measures the severity of gambling problems and classifies gamblers into four categories: non-problem gamblers (score of 0), low risk gamblers (score from 1 to 2), moderate risk gamblers (score from 3 to 7) and problem gamblers (score greater than 8).

2.2.4 Gambling Related Cognitions Scale (GRCS) (Raylu & Oei, 2004) The GRCS is a 23-item self-assessment questionnaire used to assess gambling-related cognitions. The questionnaire assesses 5 types of cognitive distortions: gambling expectancies (α=0.87), illusion of control (α=0.87), predictive control (α=0.77), inability to stop gambling (α=0.89) and interpretative bias (α=0.91). We have added 6 items to assess near-misses, others' wins, and chasing behavior (Appendix C).

2.2.5 Gambling Motives Questionnaire - Financial (GMQ-F) (Dechant, 2014; Stewart & Zack, 2008) The GMQ is a self-assessment questionnaire designed to measure the motivations behind gambling. The original version of the GMQ was constructed according to a three-dimensional model, adapted from the Drinking Motives Questionnaire (DMQ): coping, enhancement and social motives. We used a revised 15-item version of the GMQ, the GMQ-F (Devos et al., 2016), which assesses financial motives (α=0.81) in addition to coping motives (α=0.79), enhancement motives (α=0.84) and social motives (α=0.74).

2.2.6 Problematic Overuse of Poker Screen (POPS) We included a preliminary version of the POPS in the survey. The POPS was initially developed by O. Bjerg (Copenhagen Business School,

Denmark) based on a combination of several items from gambling scales which were translated into the "poker language" and then selected based on clinical relevance and questioning of poker gamblers. Some items are "specific" to poker in the sense they could not be applied to other forms of gambling; other items are "adapted" for poker in the sense they are considered relevant to poker even though they could be used for other forms of gambling. Bjerg's version included 14 items distributed according to a pre-assumed 4-dimension structure: financial, social, psychological and loss of control. To complete the assessment and as the dimensions were highly unbalanced in terms of the number of items, we added 6 new items based on items most often associated with pathological gambling or used to assess specific features of poker gamblers (e.g. poker as a mean to succeed in life, emotional stimulation) which were highlighted in a preliminary work (Bouju et al., 2013). The preliminary version of the POPS used in this study thus comprised 20 items distributed across 4 pre-assumed dimensions (5 items per dimension). The reference period was the past year.

The response categories of the 20 items of the POPS were assigned different weights. Weighting was defined so that dichotomous items (i.e. items with only 2 possible answers) had similar importance as polytomous items (i.e. items with more than 2 possible answers). We also gave more importance to answers which indicate *a priori* more severe problems related to poker playing. The 20 items of the initial version are described in Appendices A (French Version) and B (translation into English).

2.3 Statistical analysis

All analyses were performed on participants who answered at least the POPS and the PGSI. Statistical analyses were performed using Stata 14 software (College Station, TX: StataCorp LP). Quantitative variables were described as mean ± standard deviation or as median (25^{th} 75^{th} percentiles) for skewed variables. Categorical data were expressed as numbers (%). For group comparisons, qualitative variables were compared using chi-squared tests (or Fisher's exact test for sample sizes < 30) and quantitative variables were compared using Student's t-tests (or Wilcoxon-Mann-Whitney test for sample sizes < 30 and highly skewed data). Differences associated with a p-value ≤ 0.05 were considered statistically significant.

2.3.1 Primary objective

The main objective of the study was to develop a screening tool for the identification of poker-specific gambling problems. Based on the 20 items of the preliminary version of the POPS, we

aimed to achieve a medium-sized questionnaire (4–12 items) (Thomas et al., 2011) to allow for easy and fast answering.

The psychometric properties of the POPS were examined by assessing reliability, construct validity, content validity and criterion validity (Fayers & Machin, 2013). Confirmatory Factor Analysis (CFA) was used to test the structure of the 20-item preliminary version of the POPS. The following goodness of fit indices were used: Root Mean Square Error of Approximation (RMSEA) (values < 0.06 indicate good fit), Standardized Root Mean Residual (SRMR) (values < 0.06 indicate good fit) and Comparative Fit Index (CFI) (values > 0.95 indicate good fit). If all of these fit indices were found to be good, then the original structure would be validated. If that was not the case, several revised versions would be proposed based on both their psychometric properties and clinical relevance. Finally, the structure which fitted the data best was selected (structural validity).

Internal consistency was assessed by Cronbach's Alpha coefficient (Cronbach, 1951). Alpha coefficients > 0.7 were considered acceptable.

Criterion validity was assessed by looking at empirical associations between the POPS and other measures of the questionnaire related to gambling behavior (e.g. GRCS scores, control of gambling urges, damages caused by gambling). Consistent correlations between the scores of the POPS and these measures would indicate good criterion validity.

A threshold was defined to discriminate non-problem and problem poker players using the POPS. In order to define an optimal cut-off, we performed a logistic regression with the gambler's status (PGSI<8 vs. PGSI ≥8) as the dependent variable and the POPS as the independent variable. We chose to select a threshold associated with a high specificity (rate of negative test results among non-problem gamblers) to limit the number of false positives and thus not over-estimate the number of problem gamblers.

Classifications of non-problem and problem gamblers obtained with the POPS and the PGSI were finally compared quantitatively (number of concordances / non-concordances) and qualitatively (profiles of gamblers whose classification differed between the two scales).

2.3.2 Secondary objectives

Comparisons between problem and non-problem poker players according to the POPS were made in terms of sociodemographic characteristics, gambling habits, gambling motives, cognitive distortions and impulsivity profile.

3. Results

3.1 Primary objective: development of the POPS

3.1.1 Development of a medium-sized questionnaire

Items' wording are given in Appendices A and B. We decided to drop items which did not seem to be discriminative enough (i.e. not specific enough) to screen problematic poker players. As a result, we deleted item 1, item 3, item 15, item 16, item 17, item 19 and item 20 because they have high percentages of positive answers (i.e. answers different from "no" or "never" ≥ 10 %). Item 7, item 8 and item 18 were dropped based on clinical choices. With the 9 remaining items, we tested several structures based on Exploratory Factor Analysis (EFA) and clinical expertise while focusing on brief ones and we chose the structure with the best trade-off between clinical relevance and psychometric properties. The chosen solution was finally composed of 6 items: item 4, item 5, item 6, item 9, item 11 and item 12.

An EFA was performed to assess the factor structure of the 6-item questionnaire (Table 2). Rotated factor loadings suggested grouping items 4, 5 and 11 into one dimension ("negative consequences") and items 6, 9 and 12 into a second dimension ("loss of control").

[Insert Table 2 here]

3.1.2 Psychometric properties of the 6-item version

The confirmatory factor analysis showed that the 6-item structure fitted the data very well (RMSEA = 0.012, SRMR = 0.017, CFI = 0.998). The global Cronbach's alpha was 0.58.

Criterion validity was assessed by looking at the correlation coefficients between the two sub-scales of the POPS and the sub-scales of the GRCS, self-assessment of self-efficacy to control gambling urges and self-assessment of damage from gambling on different areas of life. As shown in Table 3, the "loss of control" score of the POPS is positively and strongly correlated with the GRCS-IS ("inability to stop gambling") and GRCS-chasing scores (ρ = 0.46 and 0.53, respectively), as expected. It is also negatively correlated with the ability to control gambling urges (ρ = - 0.37). The "negative consequences" score of the POPS was more strongly correlated with self-assessed damage from gambling on family, friends and job, whereas the "loss of control" score was more closely correlated with damage from gambling on health and finances.

[Insert Table 3 here]

3.1.3 Definition of a cut-off score for the POPS

We regressed the global score of the POPS against the PGSI classification (PGSI < 8 vs. PGSI ≥ 8) to define a threshold for poker overuse based on the POPS. As we wanted to find a cut-off associated with high specificity, we defined individuals with an overall POPS score ≥ 5 as excessive poker players (sensitivity = 0.52, specificity = 0.98). In our sample, 24 (3.3%) gamblers could be considered having a problematic overuse of poker according to the POPS, compared to 21 (2.9%) with the PGSI.

3.1.4 Comparison of the POPS and PGSI classifications

Table 4 shows the classification of gamblers according to the POPS and the PGSI. The scores of the GMQ-F and the GRCS and answers to two questions about gambling problems (feeling of having a gambling problem and abstinence from gambling for at least one month) were compared across the four possible groups: participants identified as non-problem gamblers by both the PGSI and the POPS (PGSI-/POPS-), participants classified as excessive gamblers by the PGSI and non-problem gamblers by the POPS (PGSI+/POPS-), participants classified as non-problem gamblers by the PGSI but problem-gamblers by the POPS (PGSI-/POPS+) and participants classified as problem gamblers by both the PGSI and the POPS (PGSI+/POPS+).

[Insert Table 4 here]

Participants of the PGSI+/POPS- group had roughly the same low negative-consequences score as participants of the PGSI-/POPS- group, who were confirmed as non-problem gamblers. On the other hand, gamblers in the PGSI-/POPS+ had high negative-consequences scores similar to those of the PGSI+/POPS+ group, who were confirmed as problem gamblers. Loss of control scores were similar between PGSI+/POPS- and PGSI-/POPS+ groups. Gamblers in the PGSI+/POPS+ group had high loss of control score.

The four GMQ scores were similar between PGSI-/POPS+ and PGSI+/POPS- gamblers. The GRCS-GE scores for PGSI-/POPS+ gamblers were higher than for PGSI+/POPS- gamblers. Moreover, even though the difference was not statistically significant, we observed higher GRCS-IS scores among PGSI+/POPS- than among PGSI-/POPS+.

All the PGSI+/POPS- participants declared to have a history of gambling-related problems but a large proportion of them declared they had problems only in the past, and only three that they

currently have problems. In comparison, PGSI-/POPS+ participants declared they had problems with gambling equally in the past and currently, although three of them declared they didn't feel as if they had a problem with gambling. The percentages of poker players who managed to stop gambling for at least one month were similar in the PGSI-/POPS-, PGSI+/POPS-, PGSI-/POPS+ groups and were lower in the PGSI+/POPS+ group.

3.2 Secondary objectives

3.2.1 Description of poker players and comparison of their characteristics according to the POPS classification

The characteristics of participants according to the POPS classification are described in Table 5. Poker players in the problematic group were younger, lived less often as a couple, started playing poker earlier, had a "big win" during their first poker experiences more often, played poker at a higher frequency, wagered more money, played aggressively more often and played a higher number of other games. As regards to gambling problems, 9 (37.5%) participants felt they had a current gambling problem and 11 (45.8%) a gambling problem in the past. Moreover, 10 out of the 20 gamblers who felt they had a gambling problem declared to have considered stopping gambling as it was harmful. Individuals in the POPS ≥ 5 group had lower scores on self-efficacy to control gambling urges. Finally, they reported greater damage in all areas of life assessed. There was a statistically significant difference ($p=0.018$) in the way participants considered poker as a game of skill or chance: 8.3% of individuals in the POPS ≥ 5 group tended to consider poker as only a game of chance against 0.6% in the POPS <5 group. Unexpectedly, there was no significant difference in the length of sessions between the two groups.

[INSERT Table 5 here]

4. Discussion

The main objective of this study was to develop a questionnaire assessing gambling problems among the specific population of poker players. The Problematic Overuse of Poker Screen (POPS) is a newly-developed, six-item, two-dimensional self-assessment questionnaire covering loss of control and negative consequences. The questionnaire globally showed acceptable psychometric properties: structural validity assessed by the confirmatory factor analysis was very good, the correlation coefficients between the POPS and sub-scales of the GRCS, self-assessment of self-efficacy to control gambling urges and self-assessment of damage of gambling on different areas of life reflected criterion validity. However, we observed a lack of internal consistency as shown by a

low Cronbach's alpha. This can be explained by the small number of items in each dimension. Indeed, as Cronbach's alpha is a function of the number of items, a small number of items can lead to small alphas values (Cortina, 1993).

We defined a cut-off score of 5 to differentiate non-problem and problem gamblers according to the POPS. We classified 24 (3.3%) participants as having problematic overuse of poker (POPS $\geq$ 5). The relatively low rate of problem gamblers may be explained by the specificity of the study population which is mainly composed of members of poker clubs. Indeed, they are more likely to play in tournaments rather than in cash games (the latter being generally the preferred form for problem players), as opposed to players who would be recruited on online poker websites for example. By comparing this classification to the PGSI classification, we found, on the one hand, that gamblers in the PGSI+/POPS- group had a negative-consequences score as low as in the PGSI-/POPS- group. On the other hand, participants in the PGSI-/POPS+ group had a negative-consequences score as high as the gamblers in the PGSI+/POPS+ group. This suggests the manifestation of gambling-related damages specific to poker that would be not adequately assessed by the PGSI. Moreover, GRCS-GE (gambling expectancies) scores were higher in the PGSI-/POPS+ group than in the PGSI+/POPS- group, which could indicate that the POPS takes gambling expectancies into account more adequately. This emotional dimension has been found to be important in poker, in particular for the distinction between non-problem and problem gamblers (Barrault et al., 2017; Bouju et al., 2013). The similar GMQ-Financial scores between PGSI+/POPS-, PGSI-/POPS+ and PGSI+/POPS+ gamblers seem to confirm that the classifications of gamblers we observed are more closely related to psychological differences rather than to financial preoccupations. In addition, Weinstock et al. (2013) showed that, compared to professional gamblers, pathological gamblers displayed poor psychological functioning, low self-efficacy and impulsivity, whereas they gambled in comparable intensities. Moreover, we observed that GRCS-IS (perceived inability to stop) scores in the PGSI+/POPS- group were higher than in PGSI-/POPS+ group (21.10 and 18.33, respectively), although not statistically different. This could indicate that the PGSI assesses the feeling of having a gambling problem more subjectively, and that PGSI-based screening for gambling problems relies more on the gambler's self-awareness regarding their potential problem with gambling. By looking at the proportions of past and current gambling problems in the PGSI+/POPS- and the PGSI-/POPS+, it may be possible that the PGSI predominantly identifies poker players with past gambling problems but fails to correctly identify players with current gambling problems, leading to a potential over-estimation of current gambling problems rates for poker players.

The secondary objective was to compare participants with a POPS score < 5 (non-problem poker gamblers) to participants with a POPS score ≥ 5 (problem poker gamblers). Besides results which seem to confirm what is generally found in the literature (problem gamblers started gambling at a younger age, often experienced a "big win" during first years of gambling, played more frequently and wagered more money (Dowling et al., 2017; Shead et al., 2008; Welte et al., 2004), we found that problem poker players declared playing aggressively "most of the time" more frequently, which could indicate a lack of flexibility and adaptation, and thus a lack of control. Siler (2010) suggested that, during a losing session, some players tend to play more aggressively trying to get even at any cost. Moreover, problem poker players showed lower scores on self-efficacy to control gambling urges, reported higher damages on all areas of life assessed and played a larger number of different games, which all together tend to support the validity of the classification with the POPS. In addition, problem poker players tended to consider poker as only a game of chance more often, compared to non-problem poker players. It has been suggested that poker players with gambling problems tend to over-estimate the luck element in poker when they are unsuccessful (Shead et al., 2008; Turner & Fritz, 2001). Unexpectedly, we found no difference in the length of poker sessions between problem and non-problem gamblers. This can be explained by the fact that problem gamblers played cash games more frequently (which often last less longer than tournaments) than non-problem gamblers.

This study is not without limitations. The participants included could represent a sub-population of poker players as they were mainly members of poker clubs. Moreover, results may suffer from self-selection bias due to the online recruitment process.

We could not assess the criterion validity of the POPS properly since we lacked a true gold standard to clinically assess gambling disorders (the PGSI being supposed to be insufficiently adapted for poker players). A future study is needed to compare the results of the POPS to DSM criteria and perhaps make adjustments to the questionnaire we developed. Regarding structural validity, a unidimensional questionnaire could have been used given EFA results; however the two dimensions are consistent with Goodman's definition of addiction (Goodman, 1990) which characterize addiction by 1) "a recurrent failure to control the behavior" (corresponding to the "loss of control" dimension) and 2) "the continuation of the behaviour despite significant negative consequences" (corresponding to the "negatives consequences" dimension). Also, the POPS should be validated in English. Furthermore, comparison of the PGSI and the POPS classifications was based on small groups of gamblers and thus the results related to these divergences in classification should be interpreted with caution.

Despite these limitations, the POPS appears to assess gambling-related damages specific to poker that would be not adequately assessed by the PGSI, and thus could contribute to fill a gap for the screening of problem poker players.

Acknowledgments

We wish to sincerely thank the “Club des Clubs” for their precious and spontaneous help in recruiting participants in this study.

This study was conducted and sponsored by Nantes University Hospital.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Conflicts of interest

MGB and GCB declare that Nantes University Hospital received funding from the gambling industry (FDJ and PMU) in the form of a sponsorship. Scientific independence with respect to these gambling industries is guaranteed and there were no constraints on publishing.

BP, JBH, OR and OB declare to have no conflicts of interest.

Ethical approval

All participants freely agreed to participate. This study was approved by the local Research Ethics Committee (GNEDS) and the anonymity of all the participants was guaranteed.

References

Association, A. P. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Pub.

Barrault, S., Bonnaire, C., & Herrmann, F. (2017). Anxiety, Depression and Emotion Regulation Among Regular Online Poker Players. *Journal of Gambling Studies*, 1‑12. https://doi.org/10.1007/s10899-017-9669-3

Barrault, S., Untas, A., & Varescon, I. (2014). Special features of poker. *International Gambling Studies*, *14*(3), 492‑504. https://doi.org/10.1080/14459795.2014.968184

Barrault, S., & Varescon, I. (2013). Impulsive sensation seeking and gambling practice among a sample of online poker players: Comparison between non pathological, problem and pathological gamblers. *Personality and Individual Differences*, *55*(5), 502‑507. https://doi.org/10.1016/j.paid.2013.04.022

Bjerg, O. (2010). Problem gambling in poker: money, rationality and control in a skill-based social game. *International Gambling Studies*, *10*(3), 239‑254. https://doi.org/10.1080/14459795.2010.520330

Bouju, G., Grall-Bronnec, M., Quistrebert-Davanne, V., Hardouin, J.-B., & Vénisse, J.-L. (2013). Texas hold'em poker: a qualitative analysis of gamblers' perceptions. *Journal of Gambling Issues*, *28*, 1‑28. https://doi.org/10.4309/jgi.2013.28.4

Boutin, C. (2010). *Le jeu: chance ou stratégie?: choisir librement la place du jeu dans votre vie*. Éditions de l'Homme.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of applied psychology*, *78*(1), 98.

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

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *psychometrika*, *16*(3), 297‑334.

Dechant, K. (2014). Show Me the Money: Incorporating Financial Motives into the Gambling Motives Questionnaire. *Journal of Gambling Studies*, *30*(4), 949‑965. https://doi.org/10.1007/s10899-013-9386-5

Devos, G., Challet-Bouju, G., Burnay, J., Maurage, P., Grall-Bronnec, M., & Billieux, J. (2016). Adaptation and validation of the Gambling Motives Questionnaire-Financial (GMQ-F) in a sample of French-speaking gamblers. *International Gambling Studies*, *0*(0), 1‑15. https://doi.org/10.1080/14459795.2016.1264080

Dowling, N. A., Merkouris, S. S., Greenwood, C. J., Oldenhof, E., Toumbourou, J. W., & Youssef, G. J. (2017). Early risk and protective factors for problem gambling: A systematic review and meta-analysis of longitudinal studies. *Clinical Psychology Review*, *51*, 109‑124. https://doi.org/10.1016/j.cpr.2016.10.008

Eroukmanoff, V., Costes, J. M., & Tovar, M. L. (2014). *Les joueurs de poker, une population présentant un profil particulier ?* Paris: Observatoire des jeux (ODJ). Consulté à l'adresse http://www.economie.gouv.fr/files/note_3_-_les_joueurs_de_poker.pdf

Fayers, P. M., & Machin, D. (2013). *Quality of life: the assessment, analysis and interpretation of patient-reported outcomes*. John Wiley & Sons.

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

Goodman, A. (1990). Addiction: definition and implications. *British Journal of Addiction*, *85*(11), 1403‑1408. https://doi.org/10.1111/j.1360-0443.1990.tb01620.x

Hopley, A. A. B., Dempsey, K., & Nicki, R. (2012). Texas Hold'em Online Poker: A Further Examination. *International Journal of Mental Health and Addiction*, *10*(4), 563‑572. https://doi.org/10.1007/s11469-011-9353-2

Hopley, A. A. B., & Nicki, R. M. (2010). Predictive factors of excessive online poker playing. *Cyberpsychology, Behavior and Social Networking*, *13*(4), 379‑385. https://doi.org/10.1089/cyber.2009.0223

Laakasuo, M., Palomäki, J., & Salmela, M. (2015). Poker players with experience and skill are not "ill"–Exposing a discrepancy in measures of problem gambling. *Journal of Gambling and Commercial Gaming Research*, *1*(1). Consulté à l'adresse http://www.gamblingstudies.org/jgcg/index.php/jgcg/article/view/83

Lesieur, H. R., & Blume, S. B. (1987). The South Oaks Gambling Screen (SOGS): a new instrument for the identification of pathological gamblers. *Am J Psychiatry*, *144*(9), 1184‑1188.

Mitrovic, D. V., & Brown, J. (2009). Poker mania and problem gambling: a study of distorted cognitions, motivation and alexithymia. *Journal of Gambling Studies*, *25*(4), 489‑502. https://doi.org/10.1007/s10899-009-9140-1

Moreau, A., Chabrol, H., & Chauchard, E. (2016). Psychopathology of Online Poker Players: Review of Literature. *Journal of Behavioral Addictions*, *5*(2), 155‑168. https://doi.org/10.1556/2006.5.2016.035

Raylu, N., & Oei, T. P. (2004). The Gambling Related Cognitions Scale (GRCS): Development, confirmatory factor validation and psychometric properties. *Addiction*, *99*(6), 757‑769.

Shead, N. W., Hodgins, D. C., & Scharf, D. (2008). Differences between Poker Players and Non-Poker-Playing Gamblers. *International Gambling Studies*, *8*(2), 167‑178. https://doi.org/10.1080/14459790802139991

Siler, K. (2010). Social and Psychological Challenges of Poker. *Journal of Gambling Studies*, *26*(3), 401‑420. https://doi.org/10.1007/s10899-009-9168-2

Stewart, S. H., & Zack, M. (2008). Development and psychometric evaluation of a three-dimensional Gambling Motives Questionnaire. *Addiction (Abingdon, England)*, *103*(7), 1110‑1117. https://doi.org/10.1111/j.1360-0443.2008.02235.x

Thomas, S., Merkouris, S., Dowling, N., Radermacher, H., Jackson, A., Misso, M., & Anderson, C. (2011). *Guideline for screening, assessment and treatment in problem gambling*. Monash University.

Turner, N. E., & Fritz, B. (2001). The Effect of Skilled Gamblers on the Success of Less Skilled Gamblers. *Journal of Gambling Issues*. https://doi.org/10.4309/jgi.2001.5.10

Weinstock, J., Massura, C. E., & Petry, N. M. (2013). Professional and pathological gamblers: similarities and differences. *Journal of Gambling Studies*, *29*(2), 205‑216. https://doi.org/10.1007/s10899-012-9308-y

Welte, J. W., Barnes, G. M., Wieczorek, W. F., Tidwell, M. C. O., & Parker, J. C. (2004). Risk factors for pathological gambling. *Addict Behav*, *29*(2), 323‑335.

Wood, R. T. A., Griffiths, M. D., & Parke, J. (2007). Acquisition, Development, and Maintenance of Online Poker Playing in a Student Sample. *CyberPsychology & Behavior*, *10*(3), 354‑361. https://doi.org/10.1089/cpb.2006.9944

Table 1: Participants' socio-demographic data (N=718)

Male	647 (90.1%)
Age	39.3 ± 11.46
Professional activity	
Employed	540 (75.2%)
Student	35 (4.9%)
Retired	41 (5.7%)
Inactive	88 (12.2%)
Other	14 (2.0%)
Living conditions	
Single	226 (31.5%)
Living as a couple	433 (60.3%)
Living with parents	41 (5.7%)
Other	18 (2.5%)
Monthly income (Euros)	1,800 [1,300-2,700]

Table 2: Rotated factor loadings for a 2-factor structure

Item	Factor 1	Factor 2
Item 4 “During the past 12 months, has your poker practice significantly disrupted your professional or educational activity?” [Au cours des 12 derniers mois, votre pratique du poker a-t-elle significativement perturbé votre activité professionnelle ou scolaire?]	0.22	0.29
Item 5 “In the past 12 months, have your gambling habits (poker and other gambling forms) caused serious problems in your relationships with family or friends?” [Au cours des 12 derniers mois, vos habitudes de jeu (poker et autres jeux) ont-elles causé de graves problèmes dans vos relations avec les membres de votre famille ou vos amis ?]	0.14	0.28
Item 11 “In the past 12 months, have you experienced such a loss at gambling (poker and other forms of gambling) that you had to borrow to manage financially?” [Au cours des 12 derniers mois, avez-vous expérimenté une telle perte d’argent au jeu (poker et autres jeux) que vous avez eu à emprunter pour vous en sortir financièrement ?]	0.29	0.36
Item 6 “In the past 12 months, have you experienced shame, guilt and / or anxiety when playing poker?” [Au cours des 12 derniers mois, avez-vous ressenti de la honte, de la culpabilité et / ou de l'anxiété lorsque vous jouez au poker ?]	0.48	0.19
Item 9 “In the past 12 months, have you ever increased your bets during a poker session by changing for significantly higher buy-ins (rush period) or by betting larger amounts on other forms of poker or gambling, such as sports betting, blackjack, roulette, slot machines and other games?” [Au cours des 12 derniers mois, vous est-il arrivé d’augmenter vos paris pendant une session de poker en changeant pour des buy-ins significativement plus élevés (période de rush) ou en pariant des montants plus importants sur d'autres types de poker ou de jeux, tels que les paris sportifs, le blackjack, la roulette, les machines à sous et d’autres jeux ?]	0.47	0.05
Item 12 “In the past 12 months, have you lost a significant amount of money in one session, in which you tried to recover even small losses by playing more uncontrollably and taking more risks than usual?” [Au cours des 12 derniers mois, avez-vous perdu un montant d’argent important en une seule session, dans laquelle vous avez tenté de recouvrer des pertes même faibles en jouant de façon plus incontrôlée et en prenant plus de risques que d'habitude ?]	0.60	0.14

Table 3: Correlation coefficients between the two dimensions of the POPS and problem-gambling related measures

	Negative consequences	Loss of control
GRCS-Gambling related expectancies	0.21	0.20
GRCS-Illusion of control	0.11	0.15
GRCS-Predictive control	0.12	0.20
GRCS-Incapacity to stop gambling	0.28	0.46
GRCS-Interpretative bias	0.22	0.30
GRCS-Others' wins	0.14	0.30
GRCS-Near misses	0.16	0.28
GRCS-Chasing	0.27	0.53
Self-assessment of self-efficacy to control gambling urges	-0.21	-0.37
Impact on family	0.32	0.28
Impact on friends	0.24	0.22
Impact on job	0.33	0.22
Impact on health	0.18	0.30
Impact on finances	0.22	0.30

Note: All Pearson's correlation coefficients are significant at the 0.01 level.

Table 4: Comparison of gamblers according to the PGSI and POPS classifications

	PGSI-/ POPS- (n=684)	PGSI+/ POPS- (n=10)	PGSI-/ POPS+ (n=13)	PGSI+/ POPS+ (n=11)
POPS-loss of control (/9)	0.52	2.50 [a]	3.08 [b]	4.82 [d]
POPS-negative consequences (/7)	0.10	0.20	2.46 [b,c]	2.64 [d]
GMQ-F Coping (/16)	5.68	7.70 [a]	8.62 [b]	9.09
GMQ-F Enhancement (/16)	10.75	11.00	12.54 [b]	12.09
GMQ-F Social (/12)	4.95	4.40	5.38	5.18
GMQ-F Financial (/16)	7.87	10.10	10.69 [b]	11.00
GRCS-GE (/28)	11.93	12.60	16.75 [b]	15.73
GRCS-IC (/28)	5.93	7.90	7.17	7.36
GRCS-PC (/42)	15.47	17.30	20.92 [b]	16.45
GRCS-IS (/35)	10.87	21.10 [a]	18.33 [b]	25.27
GRCS-IB (/28)	12.56	14.50	18.92 [b]	19.18 [d]
GRCS-Others' wins (/14)	4.42	8.10 [a]	6.42 [b]	9.55 [e]
GRCS-Near misses (/14)	5.68	8.20 [a]	9.75 [b]	7.73
GRCS-Chasing (/14)	3.64	7.20 [a]	9.58 [b]	9.91
Feeling of having a gambling problem		[a]	[b]	
No	571 (83%)	0 (0%)	3 (23%)	1 (9%)
In the past	86 (13%)	7 (70%)	5 (38%)	6 (55%)
Currently	27 (4%)	3 (30%)	5 (38%)	4 (36%)
Abstinence from gambling for at least one month				
No	42 (6%)	1 (10%)	1 (8%)	4 (36%)
Yes	435 (64%)	6 (60%)	8 (62%)	5 (45%)
Never tried	207 (30%)	3 (30%)	4 (31%)	2 (18%)

Note: GMQ-F = Gambling Motives Questionnaire-Financial; GRCS-GE = Gambling-Related Cognitions Scale-Gambling Expectancies; GRCS-IC = Gambling-Related Cognitions Scale-Illusion of Control; GRCS-PC = Gambling-Related Cognitions Scale-Predictive; GRCS-IS = Gambling-Related Cognitions Scale-Perceived Inability to Stop Gambling; GRCS-IB = Gambling-related Cognitions Scale-Interpretative Bias.

[a]Significant test between PGSI-/ POPS- and PGSI+/ POPS-

[b]Significant test between PGSI-/ POPS- and PGSI-/ POPS+

[c]Significant test between PGSI+/ POPS- and PGSI-/ POPS+

[d]Significant test between PGSI+/ POPS- and PGSI+/ POPS+

[e]Significant test between PGSI-/ POPS+ and PGSI+/ POPS+

Table 5: Comparison of non-problem and problem poker gamblers according to the POPS

	POPS < 5 (n=694)	POPS ≥ 5 (n=24)	p
Sociodemographic data			
Male	623 (89.8%)	24 (100%)	0.157
Age	40.2 ± 20.9	32.6 ± 8.6	0.003
Work situation			0.077
Employed	525 (75.7%)	15 (62.5%)	
Student	32 (4.6%)	3 (12.5%)	
Retired	41 (5.9%)	0	
Inactive	82 (11.8%)	6 (25%)	
Other	14 (2.0%)	0	
Marital situation			0.002
Single	218 (31.4)	8 (33.3)	
Couple	424 (61.1)	9 (37.5)	
Living with parents	35 (5.0)	6 (25.0)	
Other	17 (2.5)	1 (4.2)	
Monthly income (Euros)	1800 (1300–2700)	1500 (800–2350)	0.206
Gambling habits			
Introduction to gambling (age)	27.7 ± 11.5	21.2 ± 7.8	0.006
Big win during first poker experiences (yes)	184 (26.5%)	12 (50.0%)	0.011
Number of games played	4 (3-5)	5 (4–7.5)	0.009
Poker as a favorite gambling activity	653 (94.1)	22 (91.7)	0.649
Poker gambling frequency			0.018
less than once a month	16 (2.3%)	0	
once a month or more	52 (7.5%)	0	
once a week or more	349 (50.3%)	7 (29.2%)	
every day or almost every day	233 (33.6%)	12 (50.0%)	
several times a day	44 (6.3%)	5 (20.8%)	
Session length (minutes)	210 (150–270)	232.5 (142.5–300)	0.465
Monthly money wagered in poker (Euros)	50 (20–150)	300 (200–800)	<0.001
Variants played (several possible answers)			
Texas Hold'em	691 (99.6%)	24 (100%)	0.903
Omaha	220 (31.7%)	7 (29.2%)	0.793
Other	28 (4.0%)	1 (4.2%)	0.634
Cash game frequency			0.076
Never	179 (25.8%)	5 (20.8%)	
Sometimes	394 (56.8%)	10 (41.7%)	
Often	58 (8.4%)	4 (16.7%)	
Most of the time	63 (9.1%)	5 (20.8%)	
Tournament frequency			0.093
Never	17 (2.5%)	2 (8.3%)	
Sometimes	101 (14.6%)	6 (25.0%)	
Often	174 (25.1%)	6 (25.0%)	

	POPS < 5 (n=694)	POPS ≥ 5 (n=24)	p
Most of the time	402 (57.9%)	10 (41.7%)	
Do you consider poker as:			
only a game of skill	59 (8.5%)	1 (4.2%)	0.018
a game of skill and chance	631 (90.9%)	21 (87.5%)	
only a game of chance	4 (0.6%)	2 (8.3%)	
Playing aggressively (with a lot of bets and raises)			0.030
Never	52 (7.5%)	0	
Sometimes	345 (49.7%)	7 (29.2%)	
Often	204 (29.4%)	10 (41.7%)	
Most of the time	93 (13.4%)	7 (29.2%)	
Playing carefully (with few bets and raises)			0.184
Never	48 (6.9%)	3 (12.5%)	
Sometimes	294 (42.8%)	14 (58.3%)	
Often	223 (32.1%)	5 (20.8%)	
Most of the time	126 (18.2%)	2 (8.3%)	
Problem gambling history			
Feeling of having a gambling problem			
No	571 (82.3%)	4 (16.7%)	<0.001
Yes, in the past but not currently	93 (13.4%)	11 (45.8%)	
Yes, currently	30 (4.3%)	9 (37.5%)	
Considering stopping gambling because of this problem (n=143)			<0.001
No, because it would be too difficult	21 (17.1%)	3 (15.0%)	
No, because it is not really a problem for me or my family	87 (70.7%)	7 (35.0%)	
Yes, because gambling is harmful to me	15 (12.2%)	10 (50.0%)	
Having started treatment for a gambling problem (n=143)	6 (4.8)	2 (10.0)	0.310
Age at beginning of treatment (n=8)	36.5 ± 7.8	24.5 ± 31.8	1.000
Control of gambling urges (/10)	7.8 ± 2.1	5.4 ± 2.2	<0.001
Impact of gambling on family (/10)	2.5 ± 1.8	4.6 ± 2.8	<0.001
Impact of gambling on friends (/10)	2.1 ±1.7	4.1 ± 2.8	<0.001
Impact of gambling on studies/work (/10)	2.1 ± 1.8	4.7 ± 2.8	<0.001
Impact of gambling on health (/10)	2.2 ± 1.8	4.5 ± 2.8	<0.001
Impact of gambling on finances (/10)	2.0 ± 1.9	4.6 ± 2.9	<0.001

Note: Categorical variables are expressed as numbers (%), quantitative variables are expressed as mean ± standard deviation or as median (25^{th} –75^{th} percentiles);

Appendix A: Items of the 20-item initial questionnaire (French)

1.	En moyenne sur les 12 derniers mois, combien d'heures par semaine avez-vous passé à jouer au poker?	moins de 20 h = 0 entre 20 et 40 h = 1 plus de 40h = 2
2.	Au cours des 12 derniers mois, quelle a été votre perte nette au poker?	je n'ai pas perdu d'argent ou j'en ai gagné = 0 moins de 500 € = 0 plus de 500 € et moins de 3000 € = 1 plus de 3000 €=2 je ne sais pas chiffrer ma perte nette = 1
3.	Au cours des 12 derniers mois, combien de fois en moyenne vos sessions de jeu ont-elles duré plusieurs heures de plus que ce que vous aviez prévu au départ ?	jamais=0 moins d'une fois par semaine=0 1 à 2 fois par semaine=1 plus de 2 fois par semaine=3
4.	Au cours des 12 derniers mois, votre pratique du poker a-t-elle significativement perturbé votre activité professionnelle ou scolaire?	non=0 oui=2
5.	Au cours des 12 derniers mois, vos habitudes de jeu (poker et autres jeux) ont-elles causé de graves problèmes dans vos relations avec les membres de votre famille ou vos amis ?	non=0 oui=2
6.	Au cours des 12 derniers mois, avez-vous ressenti de la honte, de la culpabilité et / ou de l'anxiété lorsque vous jouez au poker?	jamais=0 parfois = 1 souvent=3
7.	Depuis les 12 derniers mois, votre sentiment de confiance en vous et / ou votre estime de vous-même dépendent-ils de si vous gagnez ou perdez au poker?	jamais=0 parfois = 1 souvent=3
8.	Au cours des 12 derniers mois, avez-vous transféré de l'argent à votre *bankroll* à plusieurs reprises au cours de la même session de poker en ligne?	jamais=0 parfois = 1 souvent=3
9.	Au cours des 12 derniers mois, vous est-il arrivé d'augmenter vos paris pendant une session de poker en changeant pour des *buy-ins* significativement plus élevés (période de rush) ou en pariant des montants plus importants sur d'autres types de poker ou de jeux, tels que les paris sportifs, le blackjack, la roulette, les machines à sous et d'autres jeux?	jamais = 0 parfois = 1 souvent = 3
10.	Au cours des 12 derniers mois, à cause du jeu (poker et autres jeux), avez-vous constitué une dette que, dans votre situation financière actuelle,	non = 0 oui=3

	vous avez du mal à payer (intérêts et/ou capital)?	
11.	Au cours des 12 derniers mois, avez-vous expérimenté une telle perte d'argent au jeu (poker et autres jeux) que vous avez eu à emprunter pour vous en sortir financièrement?	jamais = 0 parfois = 2 souvent = 3
12.	Au cours des 12 derniers mois, avez-vous perdu un montant d'argent important en une seule session, dans laquelle vous avez tenté de recouvrer des pertes même faibles en jouant de façon plus incontrôlée et en prenant plus de risques que d'habitude?	jamais = 0 parfois = 1 souvent = 3
13.	Au cours des 12 derniers mois, avez-vous volé des amis ou des membres de votre famille, ou leur avez-vous emprunté de l'argent à leur insu, dans le but de recueillir des fonds pour jouer (poker et autres jeux)?	jamais = 0 parfois = 2 souvent = 3
14.	Au cours des 12 derniers mois, avez-vous volé à d'autres personnes que des amis ou des membres de la famille, ou avez-vous commis un ou des actes illégaux, pour obtenir de l'argent pour jouer (poker et autres jeux)?	non = 0 oui =3
15.	Au cours des 12 derniers mois, le poker a-t-il été un moyen pour vous de réussir dans la vie dans le cas où votre vie personnelle et/ou professionnelle ne serait pas assez satisfaisante?	non = 0 oui = 1
16.	Au cours des 12 derniers mois, avez-vous tenté de réduire ou arrêter votre pratique du poker ?	non = 0 oui = 1
17.	Au cours des 12 derniers mois, le poker vous a-t-il permis d'atteindre un certain niveau de stimulation émotionnelle que vous ne parvenez pas ou peu à retrouver dans d'autres domaines de votre vie?	non = 0 oui = 1
18.	Au cours des 12 derniers mois, des personnes de votre entourage ont-elles critiqué votre pratique du poker?	jamais=0 parfois = 1 souvent=2
19.	Au cours des 12 derniers mois, avez-vous tenté de dissimuler la réalité de votre pratique du poker (à vos proches en particulier)?	non = 0 oui = 2
20.	Au cours des 12 derniers mois, vous êtes-vous senti envahi psychologiquement par votre	non = 0 oui = 1

	pratique du poker (par le fait d'y penser sans arrêt, de ne pas pouvoir vous concentrer sur une autre activité, d'anticiper la prochaine session, etc.)?

Appendix B: Items of the 20-item initial questionnaire (translated into English)

1.	On average, in the past 12 months, how many hours per week did you spend playing poker?	Less than 20h = 0 More than 20h but less than 40h = 1 More than 40h = 2
2.	In the past 12 months, what was your net loss in poker?	I did not lose money or I won = 0 Less than €500 = 0 More than €500 but less than €3,000 = 1 More than €3,000 = 2 I cannot quantify my net loss = 1
3.	In the past 12 months, on average, how many times did your gambling sessions last several hours longer than you originally planned?	Never=0 Less than once a week =0 Once or twice a week = 1 More than twice a week = 3
4.	During the past 12 months, has your poker practice significantly disrupted your professional or educational activity?	No = 0 Yes = 2
5.	In the past 12 months, have your gambling habits (poker and other gambling forms) caused serious problems in your relationships with family or friends?	No = 0 Yes = 2
6.	In the past 12 months, have you experienced shame, guilt and / or anxiety when playing poker?	Never=0 Sometimes= 1 Often=3
7.	For the past 12 months, have your self-confidence and / or self-esteem depended on whether you win or lose at poker?	Never=0 Sometimes = 1 Often=3
8.	In the past 12 months, have you transferred money to your bankroll several times during the same online poker session?	Never =0 Sometimes = 1 Often =3
9.	In the past 12 months, have you ever increased your bets during a poker session by changing for significantly higher buy-ins (rush period) or by betting larger amounts on other forms of poker or gambling, such as sports betting, blackjack, roulette, slot machines and other games?	Never = 0 Sometimes = 1 Often = 3
10.	In the past 12 months, because of gambling (poker and other gambling forms), have you contracted a debt that you have difficulty paying (interest and / or capital) in your current financial situation?	No = 0 Yes=3

11.	In the past 12 months, have you experienced such a loss at gambling (poker and other forms of gambling) that you had to borrow to manage financially?	Never = 0 Sometimes = 2 Often = 3
12.	In the past 12 months, have you lost a significant amount of money in one session, in which you tried to recover even small losses by playing more uncontrollably and taking more risks than usual?	Never = 0 Sometimes = 1 Often = 3
13.	In the past 12 months, have you stolen from friends or relatives, or borrowed money from them without their knowledge, in order to get money for gambling (poker and other forms of gambling)?	Never = 0 Sometimes = 2 Often = 3
14.	In the past 12 months, have you stolen from anyone other than friends or family members, or have you committed illegal acts to get money to for gambling (poker and other forms of gambling)?	No = 0 Yes =3
15.	In the past 12 months, has poker been a way for you to succeed in life in case that your personal and / or professional life would not be satisfactory enough?	No = 0 Yes = 1
16.	In the past 12 months, have you tried to reduce or stop your poker practice?	No = 0 Yes = 1
17.	In the past 12 months, has poker allowed you to reach a certain level of emotional stimulation that you cannot find in other areas of your life?	No = 0 Yes = 1
18.	In the past 12 months, have people around you criticized your poker practice?	Never =0 Sometimes = 1 Sometimes =2
19.	In the past 12 months, have you tried to hide the reality of your poker practice (especially to your loved ones)?	No = 0 Yes = 2
20.	In the past 12 months, have you felt psychologically invaded by your poker practice (by thinking about it all the time, not being able to concentrate on another activity, anticipating the next session, etc.)?	No = 0 Yes = 1

Appendix C: Items' wording of the three subscales added to the GRCS

Subscale	Item [French]
Others' wins	Knowing that somebody has won in the place where I usually gamble incites me to gamble [Savoir que quelqu'un a gagné là où je joue m'incite à jouer]
	Knowing somebody who has just won incites me to gamble [Connaître quelqu'un qui vient de gagner m'incite à jouer]
Near-misses	Having almost won makes me continue gambling [Avoir presque gagné me pousse à rejouer]
	Luck will necessarily change when I almost won [La chance va forcément tourner lorsque j'ai perdu de peu]
Chasing	I must keep gambling to get back my losses [Je dois continuer à jouer pour me refaire]
	If I lose, it is important to stick with it until I get even [Si je perds, il est important que je m'accroche jusqu'à ce que je me refasse]

Résumé de la méthodologie

La version préliminaire du questionnaire de Bjerg contenait 14 items et explorait 4 dimensions : les aspects financiers, sociaux et psychologiques, et la perte de contrôle. Afin de satisfaire la validité de contenu, nous avons ajouté six items en nous basant notamment sur les résultats d'une étude qualitative menée auprès de joueurs de poker (Bouju et al., 2013). Nous sommes donc partis d'un questionnaire composé de 20 items répartis *a priori* dans 4 dimensions (5 items par dimensions). L'objectif était d'aboutir à un questionnaire de taille moyenne (4–12 items), composé des items les plus pertinents et doté de propriétés psychométriques satisfaisantes. Après avoir défini un seuil permettant de séparer les joueurs de poker non problématiques et les joueurs de poker ayant une pratique excessive, nous avons comparé ces deux groupes sur leurs caractéristiques socio-démographiques, leurs habitudes de jeu, leurs motivations à jouer et leurs distorsions cognitives. Nous avons également comparé quantitativement et qualitativement la classification obtenue à partir du questionnaire à celle obtenue avec l'ICJE.
Le questionnaire de 20 items, ainsi que des questionnaires sur les données socio-démographiques, sur les habitudes de jeu, l'ICJE, le GRCS et le *Gambling Motives Questionnaire - Financial* (Devos et al., 2016) ont été envoyés à des joueurs membres d'associations de poker, via une enquête en ligne. Les réponses de 718 joueurs ont été analysées.

Résultats principaux

L'outil final, appelé *Problematic Overuse of Poker Screen* (POPS) était composé de 6 items explorant 2 dimensions :

- Conséquences négatives :
    - Au cours des 12 derniers mois, votre pratique du poker a-t-elle significativement perturbé votre activité professionnelle ou scolaire ?
    - Au cours des 12 derniers mois, vos habitudes de jeu (poker et autres jeux) ont-elles causé de graves problèmes dans vos relations avec les membres de votre famille ou vos amis ?
    - Au cours des 12 derniers mois, avez-vous expérimenté une telle perte d'argent au jeu (poker et autres jeux) que vous avez eu à emprunter pour vous en sortir financièrement ?
- Perte de contrôle :
    - Au cours des 12 derniers mois, avez-vous ressenti de la honte, de la culpabilité et ou de l'anxiété lorsque vous jouez au poker ?
    - Au cours des 12 derniers mois, vous est-il arrivé d'augmenter vos paris pendant une session de poker en changeant pour des *buy-ins* significativement plus élevés (période de rush) ou en pariant des montants plus importants sur d'autres types de poker ou

de jeux, tels que les paris sportifs, le blackjack, la roulette, les machines à sous et d'autres jeux ?

- Au cours des 12 derniers mois, avez-vous perdu un montant d'argent important en une seule session, dans laquelle vous avez tenté de recouvrer des pertes même faibles en jouant de façon plus incontrôlée et en prenant plus de risques que d'habitude ?

Les joueurs avec un score POPS élevé étaient plus jeunes, plus souvent célibataires, avaient commencé à jouer au poker plus jeune, avaient plus souvent expérimenté un *big win* lors de leurs premières parties de poker, jouaient plus souvent, misaient plus d'argent, jouaient de manière plus agressive (i.e. avaient tendance à jouer plus de mains et à relancer plus souvent) et jouaient à un plus grand nombre d'autres jeux. Il n'y avait en revanche pas de différence quant à la durée des sessions de poker. Par ailleurs, leurs scores d'auto-efficacité étaient plus faibles, et ils avaient plus souvent la sensation d'avoir une pratique de jeu problématique. Les résultats complets sont détaillés dans l'article.

Discussion

La structure en deux dimensions du POPS a l'avantage de correspondre à la définition de l'addiction proposée par A. Goodman (1990) : 1) un échec récurrent du contrôle du comportement (qui correspond à la dimension "perte de contrôle" du POPS) ; et 2) "la poursuite du comportement en dépit des conséquences négatives (qui correspond à la dimension "conséquences négatives"). En comparant les classifications des joueurs selon l'ICJE et le POPS, on trouvait que, d'une part, les joueurs classés comme "problématiques" selon l'ICJE (ICJE+) et "non problématiques" selon le POPS (POPS-) avaient un score de conséquences négatives moyen similaire à celui aux joueurs classés comme "non problématiques" selon l'ICJE (ICJE-) et "problématiques" selon le POPS (POPS+), et que, d'autre part, les joueurs classés ICJE-/POPS+ avaient un score de conséquences négatives moyen similaire aux joueurs classés ICJE+/POPS+. Cela suggère que les manifestations des dommages spécifiques au poker ne seraient pas évalués de manière adéquate par l'ICJE. Par ailleurs, les scores de la dimension "espoirs liés au jeu" du GRCS étaient plus élevés dans le groupe ICJE-/POPS+ que dans le groupe ICJE+/POPS-, ce qui pourrait indiquer que le POPS évalue mieux les aspects liés à la motivation. De plus, la similitude des scores de motivation financière entre les groupes PGSI+/POPS-, PGSI-/POPS+ et PGSI+/POPS+ semble suggérer que les classifications de joueurs observées sont davantage liées à des différences d'ordre psychologique plutôt qu'à des préoccupations financières. Dans ce contexte, Weinstock, Massura, et Petry (2013) ont notamment montré qu'en comparaison aux joueurs professionnels, les joueurs problématiques avaient un fonctionnement psychologique plus pauvre, de plus faibles scores d'auto-efficacité et de contrôle des impulsions, alors que leur pratique de jeu était d'une intensité comparable. Enfin, les résultats de la comparaison des proportions de joueurs avec des problèmes de jeu actuels ou passés (selon leur propre ressenti) laisse penser que l'ICJE identifierait des joueurs de poker avec des problèmes de jeu passés, mais serait moins à même d'identifier

les joueurs avec des problèmes de jeu actuels.
Concernant les propriétés psychométriques du POPS, l'analyse factorielle confirmatoire a validé la structure en deux dimensions du questionnaire. Cependant, l'analyse factorielle exploratoire effectuée en amont suggérait qu'une structure avec une seule dimension était également envisageable. L'alpha de Cronbach (Cronbach, 1951) pour l'ensemble des six items indiquait une consistance interne (0,58) en dessous du seuil généralement accepté (0,70, Nunnally Jr., 1970). Cela peut s'expliquer en partie par le faible nombre d'items (Cortina, 1993) et par le faible nombre de modalités de réponses pour certains items (en particulier les items binaires). La validité de critère – évaluée en étudiant les corrélations entre les scores du POPS et les scores du GRCS, l'auto-évaluation de d'auto-efficacité relative au contrôle des envies de jeu et des impacts sur la vie quotidienne – montrait des résultats satisfaisants. En revanche, l'accord entre le POPS et un diagnostic clinique n'a pas été mesuré. La *known-groups validity* n'a pas non plus été évaluée, de même que la reproductibilité du score. L'absence de diagnostic clinique empêchant d'évaluer correctement la validité de critère a aussi eu un impact sur la définition du seuil pour le POPS servant à distinguer les joueurs de poker non problématiques et les joueurs de poker problématiques. En effet, le seuil a été défini en essayant de maximiser la sensibilité et la spécificité du POPS en prenant comme référence l'ICJE, que l'on supposait justement insuffisamment adapté aux joueurs de poker. Par ailleurs, bien que le POPS permette de calculer deux scores (un score pour la dimension "perte de contrôle" et un score pour la dimension "conséquences négatives"), nous n'avons défini de seuil que pour le score global. Une autre limite concerne la méthode de recrutement des joueurs, sous forme d'enquête en ligne, qui a pu engendrer un biais d'auto-sélection (Khazaal et al., 2014). De plus, tous les participants étaient membres d'au moins une association de poker.
Parmi les pistes d'amélioration, le recrutement d'un autre échantillon de joueurs permettrait d'éviter certains biais, notamment celui lié au fait de développer et valider le score sur un même échantillon. Par ailleurs, il pourrait être intéressant de mettre en parallèle les scores issus du POPS et ceux d'un questionnaire récemment développé, le *Online Poker Tilt Scale* (Moreau, Delieuvin, Chabrol, & Chauchard, 2017), qui mesure la perte de contrôle liée au "tilt" (un état émotionnel ressenti par les joueurs de poker qui se traduit une déviation de leur comportement de jeu habituel et qui entraine une perte de contrôle et des pertes d'argent (Browne, 1989 ; Moreau, Delieuvin, Chauchard, & Chabrol, 2015). Enfin, la confrontation du POPS à un diagnostic posé par un professionnel de santé semble une étape nécessaire.

Chapitre 3

Évaluation des problèmes de jeu à partir de données de comportement réel

Sommaire

3.1 Données rapportées vs données de comportement réel 30

3.2 De l'analyse de données rapportées par les joueurs vers l'analyse de données de comportements de jeu réels 30

3.3 Études basées sur l'analyse de données d'activité de jeu 32

3.4 Le projet EDEIN 33

3.4.1 Motivations 33

3.4.2 Article : 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 33

3.4.3 Résumé du protocole 42

3.4.4 Difficultés liées à l'avancement du projet 42

3.4.5 Structure des données de la FDJ 43

3.4.6 Choix méthodologiques et ajustements par rapport au *study protocol* . . 44

3.1 Données rapportées vs données de comportement réel

Une des limites que l'on pourrait trouver aux questionnaires évaluant les problèmes de jeu est la précision et la véracité des réponses données par les joueurs. Ces derniers pourraient en effet rapporter des faits qui ne correspondent pas à la réalité, soit volontairement (par honte ou par déni par exemple) soit involontairement (biais de mémoire). On pourrait également penser que la discordance est encore plus grande pour les joueurs problématiques. Afin de tester ces hypothèses, Braverman, Tom, et Shaffer (2014) ont analysé la précision des estimations des joueurs quant à leurs gains et à leurs pertes, et ont comparé la précision des estimations suivant le statut problématique des joueurs (mesuré avec le *Brief Biosocial Gambling Screen* (Gebauer, LaBrie, & Shaffer, 2010). Ils ont conclu que les estimations de l'argent dépensé et gagné au cours des trois derniers mois étaient plus précises que celles concernant les douze derniers mois, et que les joueurs problématiques estimaient moins précisément leurs gains et leurs pertes. En revanche, le fait de sur-estimer ou sous-estimer l'argent dépensé ne semblait associé à aucune des variables mesurées. Dans une étude similaire, Auer et Griffiths (2017) ont montré que les joueurs étaient capables, dans une certaine mesure, d'estimer l'argent dépensé au cours du dernier mois, mais que le biais avait tendance à augmenter avec l'intensité de jeu. Les résultats suggéraient également des différences selon le type de jeu : les joueurs avaient plus de difficulté à estimer l'argent dépensé pour les jeux à gratter et les paris sportifs.
On peut noter que la pertinence des données auto-rapportées est discutée de manière plus large dans la littérature scientifique, et ce dans différents domaines. Baumeister, Vohs, et Funder (2007) ont par exemple critiqué l'abus de l'utilisation des données rapportées et le manque de données issues de l'observation des comportements réels en psychologie, dont les auteurs rappellent qu'elle se définit justement comme la science des comportements. Toujours selon Baumeister et al., le principal problème concerne la précision et la validité des données rapportées qui ne peuvent être mesurées qu'en observant en situation réelle le comportement des individus. Pour citer d'autres exemples, des discordances entre les données rapportées et les comportements réels ont été observées quant à la consommation d'alcool (Rundle-Thiele, 2009) ou encore concernant l'observance thérapeutique (Garber, Nau, Erickson, Aikens, & Lawrence, 2004).

3.2 De l'analyse de données rapportées par les joueurs vers l'analyse de données de comportements de jeu réels

Dans le contexte des JHA, les limites des données rapportées par les joueurs et surtout la nécessité de s'intéresser aux comportements de jeu réels (caractérisés par exemple par les montants d'argent misé, le temps passé à jouer, les pertes financières, etc.) ont été formalisées par Shaffer, Peller, LaPlante, Nelson, et LaBrie (2010). En effet, les auteurs sont partis du constat qu'à l'instar des comportements humains étudiées en psychologie, la description des comportements de jeu

devait passer par l'étude des comportements de jeu réels. L'étude portait plus précisément sur les comportements de jeu sur Internet, et avait pour objectifs de réaliser une revue systématique de la littérature portant sur les profils ("*patterns*") de jeu en ligne et de comparer les études basées sur les données auto-rapportées aux études basées sur l'observation de comportements de jeu. La revue systématique n'a identifié aucune étude basée sur les comportements de jeu réels jusqu'à la date du 7 mars 2008, hormis celles publiées par l'équipe de Shaffer et al. Les auteurs de l'article se sont alors davantage focalisés sur la méthodologie de ces études, et insisté sur la nécessité d'utiliser des données de comportements de jeu réels dans la recherche sur l'épidémiologie des JHA en ligne. Ils ont aussi présenté plusieurs arguments en faveur du "changement de paradigme" consistant à étudier les comportements de jeu réels au lieu d'étudier seulement les comportements auto-rapportés. Parmi les principaux arguments, on retrouvait les limites inhérentes aux données auto-rapportées (biais de mémoire, stratégies d'auto-présentation, mauvaise compréhension des questions, etc.) mais également les avantages des données reflétant les comportements de jeu réels sur Internet : la possibilité de créer des cohortes de joueurs permettant notamment d'identifier les comportements de jeu qui précèdent une pratique problématique, l'évaluation de mesures de prévention basées sur des modérateurs de jeu, et la possibilité de combiner les données objectives de comportement de jeu et celles rapportées par les joueurs pour une meilleure compréhension des déterminants du jeu problématique. Les limites des études basées sur les comportements de jeu réels étaient également discutées : les caractéristiques socio-démographiques des joueurs sont en partie auto-déclarées (et donc potentiellement inexactes), le compte d'un joueur peut en pratique être utilisé par plusieurs personnes, et les joueurs peuvent posséder un compte chez plusieurs opérateurs de jeu. Les auteurs concluaient que les futures recherches devraient intégrer les données d'observation des comportements de jeu aux données auto-déclarées relatives aux symptômes pour mettre en place des interventions de manière optimale.

Gainsbury (2011) a exprimé le même enthousiasme quant aux possibilités offertes par l'analyse des données liées aux comptes des joueurs. Selon elle, fournir aux chercheurs un accès aux données d'activité de jeu pourrait permettre d'identifier des facteurs de risque pour les problèmes de jeu et d'évaluer des mesures de jeu responsable de manière plus objective. L'article présentait aussi les avantages et les limites de la recherche sur le jeu basée sur les données liées aux comptes des joueurs. Parmi les avantages non évoqués jusqu'à présent, on trouve la possibilité d'obtenir des échantillons (voire des cohortes) de joueurs à de faibles coûts (en termes d'argent et temps), la possibilité de tester des modèles théoriques, de personnaliser les mesures de jeu responsable, et d'améliorer la compréhension conceptuelle du jeu problématique. Parmi les inconvénients, les limitées énoncées par Shaffer et al. (2010) sont reprises, mais des considérations d'ordre éthique sont également évoquées, notamment la difficulté d'obtention des consentements des joueurs dans certains cas, les mesures à mettre en place pour protéger la vie privée des joueurs le cas échéant, et le fait que les données pourrait être utilisés à des fins commerciales et allant à l'encontre de la protection des joueurs.

3.3 Études basées sur l'analyse de données d'activité de jeu

Les premières recherches basées sur l'analyse de données de comportement de jeu réel ont été produites dans le cadre d'une collaboration entre la *Division of Addictions* de la faculté de médecine de Harvard et l'opérateur de jeu en ligne Bwin. Les relations avaient été initiées par l'opérateur qui suggérait qu'il serait possible d'identifier les joueurs à risque à partir de l'observation de leurs données d'activité de jeu. Cette collaboration, ainsi que la mise à disposition des données de Bwin dans le cadre du *Transparency Project* (www.thetransparencyproject.org), ont abouti à plusieurs publications (ex : LaBrie, LaPlante, Nelson, Schumann, & Shaffer, 2007 ; LaBrie, Kaplan, LaPlante, Nelson, & Shaffer, 2008 ; Nelson et al., 2008 ; LaPlante, Schumann, LaBrie, & Shaffer, 2008 ; Broda et al., 2008 ; LaPlante, Kleschinsky, LaBrie, Nelson, & Shaffer, 2009 ; Braverman & Shaffer, 2010 ; LaBrie & Shaffer, 2011 ; Gray, LaPlante, & Shaffer, 2012 ; Braverman, LaPlante, Nelson, & Shaffer, 2013 ; Adami et al., 2013 ; LaPlante, Nelson, & Gray, 2014 ; Philander, 2014). Parmi les études ayant utilisé des données d'activité de jeu provenant d'autres opérateurs, on peut citer Dragicevic, Tsogas, et Kudic (2011) qui ont analysé les données de GTECH G2 et Luquiens et al. (2016) qui ont analysé les données de l'opérateur Winamax.
La plupart de ces études avaient pour objectif principal ou secondaire l'identification de comportements de jeu à risque, ou de joueurs à risque, ou de facteurs associés aux problèmes de jeu. Une des limites majeures que l'on retrouve systématiquement (à part dans l'étude de Luquiens et al. (2016)) concerne la manière dont était défini le statut de "joueur problématique" (ou de "joueur à risque"). En effet, différentes méthodes ont été utilisées pour approcher ce statut : 1) définir comme joueurs "atypiques" les joueurs situés au delà du 95$^{\text{ème}}$ ou du 99$^{\text{ème}}$ percentile du montant total des mises ou de nombre de paris ; ou 2) définir comme joueurs problématiques les joueurs ayant fermé leur compte en précisant que la raison était liée à des problèmes de jeu ; ou 3) utiliser des méthodes de classification non supervisée (en l'occurrence, l'algorithme des k-means) pour identifier des groupes de joueurs à risque. Dans un second temps, les données d'activité de jeu des joueurs considérés comme problématiques étaient comparées à celles des joueurs considérés comme non problématiques.
Afin notamment de pallier le problème lié à la différenciation des joueurs problématiques et non problématiques, Luquiens et al. (2016) ont recruté, via une enquête en ligne, un échantillon de joueurs de poker possédant un compte chez l'opérateur de jeu Winamax et à qui il était proposé de répondre à l'ICJE dans le cadre dans d'un projet de recherche. Par rapport aux autres études citées précédemment, celle-ci avait donc l'avantage d'identifier de manière plus précise les joueurs problématiques. Les auteurs ont ensuite comparé les données de deux groupes de joueurs (ICJE < 5 (n=11 698) vs. ICJE ≥ 5 (n=2 563)), puis ont développé un score prédictif du statut "ICJE ≥ 5". Un seuil a été défini en fixant la sensibilité à 80% ; la spécificité était de 49%. Cela signifie que 80% des joueurs problématiques étaient détectés par le modèle mais que 51% des joueurs non problématiques étaient identifiés comme des joueurs problématiques. Par ailleurs, bien que cette étude répondait à plusieurs faiblesses des autres recherches dans le domaine, et qu'elle laissait

envisager la possibilité d'implémenter un outil de repérage des joueurs problématiques utilisable en routine (sans intervention active des joueurs), elle présentait plusieurs limites : 1) le modèle intégrait uniquement les données de jeu relatives au poker et ne prenait donc pas en compte les possibles autres formes de jeu pratiquées par les joueurs sur le même site ; 2) l'activité de jeu sur d'autres opérateurs n'avait pas pu être prise en compte ; 3) le modèle pourrait ne pas être généralisable aux formes de jeu autres que le poker ; 4) les résultats du modèle n'avaient pas été confrontés à des données cliniques.

3.4 Le projet EDEIN

3.4.1 Motivations

En 2013, l'ARJEL a rédigé un rapport développant un ensemble de propositions opérationnelles pour lutter contre le jeu excessif et pathologique (ARJEL, 2013). Une de ces propositions consistait à "imposer aux opérateurs de mettre en place un dispositif de repérage des joueurs des joueurs problématiques et pathologiques sur la base d'indicateurs encadrés par l'ARJEL". Dans l'optique de répondre à ce besoin d'outil de repérage des joueurs problématiques, un projet de recherche porté par des membres de l'Institut Fédératif des Addictions Comportementales (CHU Nantes) et de l'UMR INSERM 1246 (Université de Nantes, Université de Tours, INSERM) a été soumis dans le cadre d'un appel à projets de l'Institut de Recherche en Santé Publique et de ses partenaires. Le projet, nommé "EDEIN" pour "Etude de Dépistage des comportements Excessifs de jeu sur INternet", a également pour ambition de répondre aux limites des études présentées dans la section 3.3. Le protocole de l'étude est présenté dans l'article ci-dessous, suivi d'un résumé en Français.

3.4.2 Article : 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

Downloaded from http://bmjopen.bmj.com/ on March 13, 2018 - Published by group.bmj.com

BMJ Open

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

Bastien Perrot,[1,2] Jean-Benoit Hardouin,[1] Jean-Michel Costes,[3] Julie Caillon,[1,2] Marie Grall-Bronnec,[1,2] Gaëlle Challet-Bouju[1,2]

To cite: Perrot B, Hardouin J-B, Costes J-M, *et al*. 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* 2017;**7**:e014600. doi:10.1136/bmjopen-2016-014600

► Prepublication history and additional material are available. To view these files please visit the journal online (http://dx.doi.org/ 10.1136/bmjopen-2016-014600).

Received 6 October 2016
Revised 14 February 2017
Accepted 23 March 2017

[1]Université de Nantes, Université de Tours, INSERM, SPHERE U1246, Nantes, France
[2]Department of Addictology and Psychiatry, CHU Nantes, Clinical Investigation Unit 'Behavioral Addictions/Complex Affective Disorders', Nantes, France
[3]Observatoire des jeux, Ministère des Finances, Paris, France

Correspondence to
Bastien Perrot;
bastien.perrot@univ-nantes.fr

ABSTRACT

Introduction Since the legalisation of online gambling in France in 2010, gambling operators must implement responsible gambling measures to prevent excessive gambling practices. However, actually there is no screening procedure for identifying problematic gamblers. Although several studies have already been performed using several data sets from online gambling operators, the authors deplored several methodological and clinical limits that prevent scientifically validating the existence of problematic gambling behaviour. The aim of this study is to develop a model for screening excessive gambling practices based on the gambling behaviours observed on French gambling websites, coupled with a clinical validation.

Methods and analysis The research is divided into three successive stages. All analyses will be performed for each major type of authorised online gambling in France. The first stage aims at defining a typology of users of French authorised gambling websites based on their gambling behaviour. This analysis will be based on data from the Authority for Regulating Online Gambling (ARJEL) and the Française Des Jeux (FDJ). For the second stage aiming at determining a score to predict whether a gambler is problematic or not, we will cross answers from the Canadian Problem Gambling Index with real gambling data. The objective of the third stage is to clinically validate the score previously developed. Results from the screening model will be compared (using sensitivity, specificity, area under the curve, and positive and negative predictive values) with the diagnosis obtained with a telephone clinical interview, including diagnostic criteria for gambling addiction.

Ethics and dissemination This study was approved by the local Research Ethics Committee (GNEDS) on 25 March 2015. Results will be presented in national and international conferences, submitted to peer-reviewed journals and will be part of a PhD thesis. A final report with the study results will be presented to the ARJEL, especially the final screening model.

Trial registration number NCT02415296.

Strengths and limitations of this study

- This study will contribute to setting up an innovative prevention measure used to inform and protect gamblers as early as possible.
- Data will be representative of the French online gamblers' population; all types of authorised online gambling forms, and data from all authorised operators in France will be included.
- The final screening model will be clinically validated.
- We will use advanced statistical methods to build the screening model.
- Selection bias may occur during participants' recruitment at stages 2 and 3.
- Owing to technical and confidentiality constraints, some potentially interesting gambling indicators will not be used.

INTRODUCTION

The majority of gamblers have a controlled and recreational gambling practice, but some of them lose control of it. In the Fifth Edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5), gambling disorder is defined as a 'persistent and recurrent problematic gambling behavior leading to clinically significant impairment or distress'.[1] A French national survey conducted in 2015 indicated that 3.9% of past-year gamblers had gambling problems, including 0.9% of excessive gamblers.[2] With regard to online gambling only, Tovar *et al*[3] found a prevalence of gambling problems of 17% among Internet gamblers, including 6.6% of excessive gamblers. Similar rates were found by Wood and Williams[4] among a sample of Canadian gamblers; indeed, the authors of this study found that 5.7% of non-Internet gamblers could be considered as problem gamblers, including 1.7% of excessive gamblers, compared with a prevalence of gambling problems of 16.6% among Internet gamblers, including 3.8%

Downloaded from http://bmjopen.bmj.com/ on March 13, 2018 - Published by group.bmj.com

of excessive gamblers. Several factors may explain the fact that online gambling is possibly more addictive than offline gambling, especially greater accessibility, increased disinhibition and higher event frequency.[5]

The opening to competition and regulation of the French online gambling sector was provided by the law No 2010–476 of 12 May 2010. This law included the creation of the Regulatory Authority for Online Gambling (ARJEL in French). The ARJEL can issue licences to online operators for only three types of games: horse race betting, sports betting and poker. Lotteries and scratch games provided by the Française des Jeux (FDJ, ie, France's national lottery operator) are also authorised online because of a particular waiver. The ARJEL compiles data from all accounts created on authorised online gambling sites in France, namely 1352 000 active accounts in the first quarter of 2016.

In 2013, the ARJEL drafted a report including 33 proposals to fight against excessive online gambling.[6] One of them concerned the introduction of a system for identifying problematic and disordered gamblers based on indicators available in the ARJEL database. The aim of this study is to contribute to this proposal by developing a model for the identification of excessive gamblers and extend it to all authorised gambling types.

Several studies have already been performed using a database from a single online gambling operator: data from Bwin were used by LaBrie *et al*[7] LaBrie and Shaffer,[8] Broda *et al*[9] Philander[10] and Adami *et al*;[11] data from GTECH G2 were used by Dragicevic *et al*[12] and data from Winamax were used by Luquiens *et al*[13]

In most of them, one of the objectives was to identify potential problem gamblers and/or find indicators or behaviours associated with problem gambling. As the gamblers' status (eg, non-problem or problem gambler) were a priori unknown, two types of methodology had been considered. The first method, applied by Dragicevic *et al*[12] and Adami *et al*[11] was to cluster the gamblers into homogenous classes. Then, by interpreting the classes, the authors could identify groups of gamblers with potentially risky behaviours. However, as highlighted by Adami *et al*[11] 'it is difficult to assign any of the clusters to specific clinical groups with a high degree of certainty with respect to problem gambling'. The second solution was to approximate the gamblers' status by a proxy. In this case, the chosen proxy served as a reference to compare groups or predict the value of the proxy by using gambling data. For example, LaBrie *et al*[7] compared the top 1% most involved gamblers (in terms of amount wagered or number of bets) with the remaining 99%. LaBrie *et al*[8] and Philander[10] compared the behaviour of gamblers who closed their accounts because of gambling-related problems with the behaviour of other account closers. The relevance of these various proxies is questionable, and, as stated by,[11] a solution could be 'integrating gambling data with psychological studies using structured interviews in order to determine the level of gambling problems'.

In a recent study using the Winamax database,[13] the authors collected actual gambling data from a sample of poker players as well as responses to the Problem Gambling Severity Index (PGSI).[14] Gamblers were separated into a non-problem gamblers group (PGSI<5) and a problem gamblers group (PGSI≥5). A multivariate logistic regression was performed on gambling data to estimate the probability that an individual was screened as a problem gambler. However, even if the PGSI is often considered as a good screening tool, the clinical diagnosis based on the DSM-5 criteria remains the gold standard in assessing gambling disorders.

With regard to the results of this study, the specificity of the model developed was only 49.3%. Although the authors argue that 75% of 'false-positive gamblers' (ie, the gamblers wrongly classified as problem gamblers, 50.7%) had responded positively to at least one question on the PGSI, a more discriminant model might have been achieved by using another type of algorithm. In particular, the authors explain that quantitative variables were categorised into quartiles, which lead to a loss of information compared with treating them as continuous. Furthermore, in a prediction perspective, it is often worth to train multiple models and select the best one. Moreover, like most of the studies using real gambling data, this work focused on only one type of gambling from a single operator, whose clients may not be representative of all online gamblers in a given country.

In conclusion, several studies have been previously performed to identify problem online gambling, but they displayed multiple limits. The EDEIN (Screening for Excessive Gambling Behaviors on the Internet) study aims at overcoming these weaknesses by proposing a model for an early prediction of online gambling problems based on player account-based gambling data, applicable to all types of gambling and clinically validated. Our objective is also to identify distinct gamblers' profiles based on their gambling behaviour in order to set up targeted prevention measures.

METHODS AND ANALYSIS

The general objective of the study is to develop a model for screening excessive gambling practices based on the behaviour of online gamblers. The research is split into three successive stages.

First stage

The first stage aims at defining a typology of French online gamblers based on their player account-based gambling data. The objective is to identify classes of gamblers with gambling behaviours that are potentially indicative of excessive gambling.

The study population will be an anonymised random sample of 20 000 users of French gambling websites certified by the ARJEL (n=10 000) and from the FDJ (n=10 000). Only validated accounts for which the gambler has placed at least one bet during the inclusion

Downloaded from http://bmjopen.bmj.com/ on March 13, 2018 - Published by group.bmj.com

period will be included. The set of available variables will include sociodemographic data (age and sex), data concerning the accounts (eg, modification of bet limits, number of money deposits) and data related to gambling itself (eg, number of active betting days, amount of wagers, net loss). Despite their potential interest for screening excessive gambling practice, some indicators cannot be extracted due to technical (eg, poker multitabling, duration of sessions, chasing) or confidentiality constraints (eg, amount of each deposit and bet). For several indicators, we will also measure their evolution between the three previous months and the last 30 days. This will allow us to assess intragambler variability and detect individuals who changed their gambling habits early, especially risk-taking changes (eg, increase in net loss or number of money deposits). We will use these indicators to define a typology of online gamblers. Individuals with similar gambling practices (eg, similar gambling frequencies, similar number of deposits per month, etc) will be grouped in the same class. By interpreting the classes, we might be able to find groups of presumably risky gamblers and identify the indicators most related to potential gambling problems.

The classification process will be based on the monthly data of gamblers. The statistical unit will be the data of a gambler observed over the last 30 days and, for certain indicators (eg, total stake, total number of bets, total deposit, etc), a comparison between the given month and the three previous months. Accordingly, to classify a gambler at month *m*, we will use his gambling activity during month *m* plus the evolution (expressed as a difference or a ratio according to the indicators) of gambling activity between *m* and the average gambling activity measured at *m-3*, *m-2* and *m-1*. Thus, as we need a 3-month 'follow-up' for each monthly classification, gamblers will not be classified for the first 3 months of the inclusion period. The same gambler will not necessarily be assigned to the same class for the 9 months.

In order to take into account the specificity of each major type of gambling, five clustering models will be developed. Five sets of variables will therefore be taken into account (table 1): set 1 is composed of variables that are 'independent' of the type of game (eg, money deposit, gambling limits), set 2 consists of variables which can be computed for each of the four types of games and can be used either for the global gambling practice by summing all types of games or for each game separately (eg, money wagered, use of bonus), set 3 is composed of specific sports betting variables (eg, live betting, complex sports bets), set 4 contains horse race betting variables (eg, complex horse race bets) and set 5 is composed of specific lottery and scratch game variables (eg, differed lottery). There were no specific poker variables available.

The five sets of variables will be used to define five models (table 1). The first model ('global status') will be based on variables of set 1 and variables of set 2 computed for the four types of gambling together; it represents the global activity of the gambler, irrespective of the gambling type. The second model ('sports betting') will be based on variables from set 1, variables from set 2 computed for sports betting and variables from set 3; it represents the gambling activity for sports betting only. The third model ('horse race betting') will use variables from set 1, variables from set 2 computed for horse race betting and indicators from set 4; it represents the gambling activity for horse race betting only. The fourth model will be based on variables from set 1 and variables from set 2 computed for poker; it represents the gambling activity for poker only. The fifth model will use variables from set 1, variables from set 2 for lotteries and scratch cards and variables from set 5; it represents the gambling activity for lotteries and scratch card games only. As a result, we will obtain five classifications. For example, if a gambler who plays multiple games is in the cluster composed of potential problem gamblers in the global classification, it will be interesting to look at which cluster he/she belongs to in the game-specific classifications. This strategy could lead to more preventive actions targeted towards the problematic games only.

The statistical method used for the classification of gamblers will be a latent class clustering analysis.[15] In this model-based clustering approach, we assume that the data are generated from a mixture of underlying probability distributions. In other words, we assume that the population of gamblers observed is heterogeneous, composed from several but a priori unknown homogenous subpopulations (eg, casual gamblers, risky gamblers, etc). One of the advantages of latent class clustering is the possibility to compute statistical criteria based on the likelihood of the model (eg, Akaike Information Criterion and Bayesian Information Criterion) or accuracy of the classification (entropy) in order to choose the number of clusters and evaluate the fit of the models. This approach also allows to compare various model specifications (for the same number of classes) by adding or relaxing some constraints as described in.[16] These constraints are, for instance, considering that the variance of (continuous) observed variables is invariant in the different classes or assuming that observed variables are independent in each cluster (local independence). Adding such constraints allows to reduce the number of parameters of a latent class model which can grow rapidly with the number of classes. Moreover, latent class models allow to estimate two types of probabilities: the probability to observe an individual's characteristic given the latent class and the probability that an individual belongs to a latent class given his/her characteristics. For example, we can estimate the probability that a gambler in class *k* makes two deposits of money in the month and the probability that a gambler who makes two deposits of money belongs to class *k*. Another advantage of latent class clustering is the possibility of including both continuous and categorical variables via the use of appropriate distribution (typically Gaussian distributions for continuous indicators and binomial/multinomial distributions for binary/ordinal indicators).

Downloaded from http://bmjopen.bmj.com/ on March 13, 2018 - Published by group.bmj.com

Table 1 Sets of variables used to define the five models

	Set 1 (variables 'independent' from the type of game)	Set 2 (variables computed for each type of game)	Set 3 (specific sports betting variables)	Set 4 (specific horse race betting variables)	Set 5 (variables specific to lotteries and scratch games)
	► Total deposit ► Total number of deposits ► Biggest deposit in a single day ► Total withdrawal ► Number of sequences of three deposits within a 12-hour period (chasing 1) ► Number of times when we observe a deposit made less than 1 hour after a bet (chasing 2) ► Number of different types of games played ► Number of changes in the wagering limits ► Number of changes in the deposit limits ► Number of changes in auto-withdrawal limits ► Highest limit set in wagering limits ► Highest limit set in deposit limits ► Number of active accounts	► Total stake ► Total number of bets ► Net loss ► Total win in the previous month ► Number of gambling days ► Coefficient of variation of the number of bets for gambling days ► Coefficient of variation of stakes for gambling days ► Biggest total stake in a single day ► Total bonuses used	► Total stake for complex bets ► Total stake for live bets ► Number of different sports	► Total stake for complex bets	► Number of different games played ► Type of lottery (instant/ deferred)
Global status	✓	✓ (For all types of game)			
Sports betting	✓	✓ (For sports betting)	✓		
Horse race betting	✓	✓ (For horse race betting)		✓	
Poker	✓	✓ (For poker)			
Lotteries and scratch games	✓	✓ (For lotteries and scratch games)			✓

Downloaded from http://bmjopen.bmj.com/ on March 13, 2018 - Published by group.bmj.com

After estimating the model parameters, groups of gamblers will be interpreted by describing the variables in each class by their mean, median or category probabilities according to the nature of the variable. Differences between clusters for each indicator will be assessed by global and pairwise tests.

Second stage

The objective of the second stage is to define a score in order to predict whether a gambler is problematic or not. The score will be obtained by crossing the player account-based gambling data with the results of the PGSI, obtained from an online questionnaire.

The PGSI will be issued to a panel of active online gamblers who agree to reply voluntarily and anonymously. Every gambler will therefore be assigned a status depending on the results of the PGSI (high risk of gambling problems, moderate risk, low risk or no risk). The gambling data of the respondents will be linked to the answers to the questionnaire thanks to the use of an encrypted identifier.

Based on an estimated 1%–3% response rate, we expect about 20 000 gamblers to answer the questionnaire (10 000 for the ARJEL and 10 000 for the FDJ). Based on their score on the PGSI, gamblers will be classified in one of the four following categories: score of 0: no gambling problem; score of 1 or 2: low risk of gambling problem; score of 3 to 7: moderate risk of gambling problem; score of 8 or more: excessive gambling problem.

In order to predict gamblers' status defined by the PGSI, several supervised learning algorithms could be applied: for example, logistic regression, support vector machines (SVM)[17] and random forest.[18] A recent article[10] compared the performance of several data mining procedures to identify high-risk online sports gamblers (individuals who closed their account due to gambling-related problems). In particular, logistic regression, LASSO regression,[19] artificial neural networks (ANN),[20] SVM and random forests were compared in terms of sensitivity, specificity, accuracy, precision and area under the curve (AUC). The results show that none of the algorithms tested gave acceptable results (the sensitivity ranged from to 1.8% to 29.1%). ANN was the method who performed the best in the hold-out sample with a sensitivity of 29.1% and a specificity of 81.1%. In addition to the comparison of statistical methods, one of the conclusions was that variables included in the model (ie, variables from the Bwin database used in[21] and demographic variables) were insufficient to correctly predict problem gamblers. The author highlighted the need to identify new behavioural variables to build more efficient models. In our project, we attempt to overcome these weaknesses by integrating such behavioural variables (eg, chasing proxies or behaviour changes over time), and by testing several supervised learning algorithms in order to select the most efficient one.

We will measure the sensitivity (probability to detect gamblers at risk), specificity (probability to detect gamblers not at risk), positive predictive value (probability that an individual detected at risk is really at risk), negative predictive value (probability that a gambler detected not at risk is really not at risk) and AUC for each method used. Other measurements, such as Brier's scores and calibration assessment, could be computed depending on the algorithm used. In order to limit overfitting, data will be partitioned into learning sample, validation sample and test sample or/and cross-validation will be applied, depending on the method used.

Third stage

The third stage of the study aims at clinically validating the screening model obtained at the second stage, by comparing the predictions of the model with current diagnosis of gambling disorder based on the National Opinion Research Center DSM-IV Screen for Gambling Problems (NODS).[22] We will use a revised version of the NODS that we have created to take into account the changes in the gambling disorders section in the DSM-5. Furthermore, we will diagnose lifetime and past 12 months gambling disorders. Individuals with current gambling problems will be asked about the presence of symptoms for each of the past 12 months.

Participation in stage 3 will be proposed after completing the online questionnaire at stage 2. We will also propose participation in stage 3 to gamblers registered in our clinical unit's volunteer base. Once eligibility is confirmed, and if the volunteer actually accepts to take part in stage 3, a telephone clinical interview conducted by well-trained staff members with experience with pathological gamblers will be offered, lasting about 30 min. It will make it possible to define the gambler's clinical status (presence or absence of a gambling disorder diagnosis), according to the DSM-5 definition. The interview will also comprise a set of questions on the gambling course and habits, motivations (assessed by the Gambling Motives Questionnaire-Financial[23]), gambling-related cognitions (assessed by the Gambling Related Cognitions Scale)[24 25] and negative consequences (assessed by questions measuring the impact of gambling on different areas of life and a scale currently being validated). Participants will be given a €50 gift voucher in compensation for their participation.

We expect at least 240 participants (60 for each type of gambling), including half with a gambling problem detected by the PGSI (threshold of 8) and half with no gambling problem. The total number of participants was computed so that we obtain at least 30 individuals by group. The diagnosis obtained from the telephone clinical interview will be crossed with the screening model results to assess the clinical validity of the model. As for stages 1 and 2, analyses will be carried out separately between the principal four types of gambling: poker, sports betting, horse race betting, lotteries and scratch games. This third stage will also eventually allow the identification of new or improved proxies of risky gambling behaviour, especially based on time indicators (which are

Downloaded from http://bmjopen.bmj.com/ on March 13, 2018 - Published by group.bmj.com

Table 2 Summary of data collected throughout the three phases

Phase I	Phase II	Phase III
► Gambling data from a random panel of gamblers' accounts extracted from the ARJEL and FDJ databases (ie, variables from table 1)	► Data obtained from the online questionnaire ► Gambling data of participants to phase II (new data from the ARJEL and FDJ databases)	► Sociodemographic data ► Variables on gambling habits ► Evaluation of cognitive distortions ► Diagnosis of gambling disorders ► Gambling data obtained from participants' account history (especially time-related variables) ► Gambling data of participants to phase III (new ARJEL and FDJ data extraction)

ARJEL, Authority for Regulating Online Gambling; FDJ, Française Des Jeux.

not available in the data sets due to difficulties to extract them; eg, duration of gambling per day, chasing, etc). A summary of data collected through the three stages of the project is shown in table 2.

DISCUSSION

The main objective of this study is to develop a model for screening problematic online gambling behaviours using player account data. Gambling disorders will be assessed by the PGSI and results will be crossed with gamblers' data and validated clinically with DSM-5-based diagnoses. Latent class clustering and supervised learning algorithms will be used to estimate several models and choose the best one.

Since all online gambling operators authorised in France will be involved, data will be representative of the French gamblers' population, overcoming the limitations of previous studies based on data from a single gambling operator. This will increase the generalisability of the results and enable routine use of the screening model on any gambling website. Another strength of this project is the clinical validation of the screening model assessed through a clinical interview of a subset of gamblers. This is the first time that a diagnosis-based clinical validity (rather than screening via an autoquestionnaire) is used for this kind of study, thus bringing a high construct validity to the developed model. This study also takes into account the specificity of gambling types, which allows the detection of gambling problems at an individual level (with the overall gambling practice, whatever the gambling type) and separately for the different types of online gambling (poker, sports betting, horse race betting, lotteries and scratch cards).

Moreover, variables used in stages 1 and 2 are not just raw data but rather have been defined specifically to be related to risky gambling behaviours.

In the first stage, latent class models may produce different results compared with more traditional methods, like k-means or hierarchical clustering, used in the context of online gambling data. In particular, these latter methods sometimes generate several very homogenous clusters and a cluster grouping all extreme profiles. However, problem gambling behaviour is considered as one particular extreme behaviour that we want to identify precisely. The use of latent variable models can help to detect this kind of behaviour. In the second stage, the use of different kinds of supervised algorithms may allow the identification of the best-performing model(s) in order to predict individuals with gambling problems. Moreover, it will be interesting to compare the clusters obtained from the latent class analysis with the screening model prediction.

Some limitations of this study include potential self-selection bias[26] at stages 2 and 3. To uncover these biases, player account-based gambling data will be compared between individuals who completed the PGSI (stage 2) and a representative sample of non-selected gamblers (stage 1), as well as between gamblers interviewed (stage 3) and those not interviewed during the clinical phase (stages 1 and 2). With regard to the PGSI, the cut-offs for the low-risk and moderate-risk categories have been criticised in the literature and alternative cut-offs have been proposed to produce more discriminant intermediate categories.[27] It would be interesting to perform sensitivity analyses according to the definition of the two intermediate thresholds. Furthermore, the collected data will be representative of users of French authorised gambling websites but we will have no information about users of unlicensed gambling websites. A study by Costes *et al*[28] showed that 46.3% of online gamblers gambled on at least one unlicensed site, of which 12.1% gambled exclusively on unlicensed sites. As another limitation, player account-based gambling data will not include time-based indicators, because of the complexity of their calculation on a large scale. This limit has been highlighted in previous studies as a potential bias.[13] Moreover, building new behavioural variables was recommended in order to develop more efficient models.[10] Thanks to the third clinical stage of the study, we will collect more detailed data (especially time-related data) and may be able to find new proxies of risky gambling behaviours, such as chasing behaviour or time spent gambling, by gaining access to the gamblers' account history during clinical interviews.

CONCLUSION

Thus, the results of this study may lead to the implementation of a system for the early identification of problem

Downloaded from http://bmjopen.bmj.com/ on March 13, 2018 - Published by group.bmj.com

and disordered gamblers based on player account-based gambling data. The project will therefore contribute to setting up an innovative and effective prevention measure, in order to inform and protect gamblers as early as possible. In particular, the need for prevention tools has been highlighted for Internet gamblers,[29] and the implementation of gambling moderators has been encouraged.[30 31] This tool will have the advantage of being quasiroutinely usable on any gambling website since no intervention from the gamblers will be required to define their status.

Specific information and advice could be provided early for individuals identified as at-risk or problem gamblers. In addition to the screening of gambling problems, the identification of distinct gamblers' profiles based on their gambling behaviour (eg, individuals gambling a lot of money during short periods, gamblers with varying gambling behaviours, individuals playing many different games, etc) will allow the implementation of targeted prevention measures. This includes messages focused on time spent gambling for gamblers with limited financial damage and high familial and social damage, information on the randomness of outcomes for gamblers with a high level of cognitive distortions or recommendation of pauses during the game for gamblers with an important chasing behaviour. Moreover, a final report with the study results, especially the final screening model, will be presented to the ARJEL in reference to its report to combat excessive online gambling.[6]

Acknowledgements We would like to warmly thank the ARJEL and FDJ for their contribution to this project, by allowing us to have access to the gamblers' data. This research is conducted on the initiative of and coordinated by the Clinical Investigation Unit BALANCED 'BehaviorAL AddictioNs and ComplEx mood Disorders' of the University Hospital of Nantes, which sponsors this study.

Contributors BP conducted the literature research on statistical methods and will conduct the statistical analysis. J-BH provided methodological advice, designed the statistical analysis plan and will supervise the statistical analysis. J-MC and GC-B selected the parameters to be included in the study. J-MC designed and was responsible of the questionnaire issued in stage 2 to ARJEL gamblers. JC provided Internet gambling and prevention advice. MG-B is the principal investigator and provides the study's medical supervision. GC-B designed the study, wrote the protocol and is in charge of project management. BP wrote the first draft of the manuscript and all authors read and approved the final manuscript.

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 Behaviors (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.

Competing interests MG-B, JC and GC-B declare that the University Hospital of Nantes has received gambling industry (FDJ and PMU) funding in the form of a sponsorship which 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. BP, J-MC and J-BH declare that they have no conflict of interest

Patient consent For stage 2 and 3, participants were informed about the research. For stage 3, they gave their written informed consent prior to their inclusion in the study.

Provenance and peer review Not commissioned; externally peer reviewed.

Open Access This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2017. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES

1. Association AP. Diagnostic and statistical manual of mental disorders (DSM-5®). *American Psychiatric Pub* 2013.
2. Costes J-M, Eroukmanoff V, Richard J-B, *et al*. Les jeux d'argent et de hasard en France en 2014. *Notes ODJ N°* 2015;6.
3. Tovar ML, Costes JM, Eroukmanoff V. Les jeux d'argent et de hasard sur internet en France en 2012. *Tendances* 2013;85:1–6.
4. Wood RT, Williams RJ. A comparative profile of the internet gambler: demographic characteristics, game-play patterns, and problem gambling status. *New Media Soc* 2011;13:1123–41.
5. Griffiths M. Internet gambling: issues, concerns, and recommendations. *Cyberpsychol Behav* 2003;6:557–68.
6. ARJEL. Lutter Contre le jeu excessif ou pathologique. recommandations trois ans après l'adoption de la loi d'ouverture du marché des jeux en ligne. http://www.arjel.fr/IMG/pdf/20130426-addiction.pdf
7. LaBrie RA, LaPlante DA, Nelson SE, *et al*. Assessing the playing field: a prospective longitudinal study of internet sports gambling behavior. *J Gambl Stud* 2007;23:347–62.
8. LaBrie R, Shaffer HJ. Identifying behavioral markers of disordered internet sports gambling. *Addict Res Theory* 2011;19:56–65.
9. Broda A, LaPlante DA, Nelson SE, *et al*. Virtual harm reduction efforts for internet gambling: effects of deposit limits on actual internet sports gambling behavior. *Harm Reduct J* 2008;5:27.
10. Philander KS. Identifying high-risk online gamblers: a comparison of data mining procedures. *Int Gambl Stud* 2014;14:53–63.
11. Adami N, Benini S, Boschetti A, *et al*. Markers of unsustainable gambling for early detection of at-risk online gamblers. *Int Gambl Stud* 2013;13:188–204.
12. Dragicevic S, Tsogas G, Kudic A. Analysis of casino online gambling data in relation to behavioural risk markers for high-risk gambling and player protection. *Int Gambl Stud* 2011;11:377–91.
13. Luquiens A, Tanguy ML, Benyamina A, *et al*. Tracking online poker problem gamblers with player account-based gambling data only. *Int J Methods Psychiatr Res* 2016;25:333–42.
14. Ferris J, Wynne H. The canadian problem gambling index. *Ott Can Cent Subst Abuse* 2001.
15. Vermunt JK, Magidson J. Latent class cluster analysis. *Appl Latent Cl Anal* 2002:89–106.
16. Fraley C, Raftery AE. How many clusters? which clustering method? answers via model-based cluster analysis. *Computer J* 1998;41:578–88.
17. Cortes C, Vapnik V. Support-vector networks. *Mach Learn* 1995;20:273–97.
18. Breiman L. Random forests. *Mach Learn* 2001;45:5–32.
19. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B Methodol* 1996:267–88.
20. Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychol Rev* 1958;65:386–408.
21. Braverman J, Shaffer HJ. How do gamblers start gambling: identifying behavioural markers for high-risk internet gambling. *Eur J Public Health* 2012;22.
22. Wickwire EM, Burke RS, Brown SA, *et al*. Psychometric evaluation of the National Opinion Research Center DSM-IV Screen for Gambling problems (NODS). *Am J Addict* 2008;17:392–5.
23. Devos G, Challet-Bouju G, Burnay J, *et al*. Adaptation and validation of the gambling motives Questionnaire-Financial (GMQ-F) in a sample of French-speaking gamblers. *Int Gambl Stud* 2016:1–15.
24. Raylu N, Oei TP. The Gambling Related Cognitions Scale (GRCS): development, confirmatory factor validation and psychometric properties. *Addiction* 2004;99:757–69.
25. Grall-Bronnec M, Bouju G, Sébille-Rivain V, *et al*. A french adaptation of the Gambling-Related Cognitions Scale (GRCS): a useful tool for assessment of irrational thoughts among gamblers. *Journal of Gambling Issues* 2012;27:1–21.

Downloaded from http://bmjopen.bmj.com/ on March 13, 2018 - Published by group.bmj.com

26. Khazaal Y, van Singer M, Chatton A, *et al*. Does self-selection affect samples' representativeness in online surveys? an investigation in online video game research. *J Med Internet Res* 2014;16:e164.
27. Currie SR, Hodgins DC, Casey DM. Validity of the Problem Gambling Severity Index interpretive categories. *J Gambl Stud* 2013;29:311–27.
28. Costes JM, Kairouz S, Eroukmanoff V, *et al*. Gambling patterns and problems of gamblers on licensed and unlicensed sites in France. *J Gambl Stud* 2016;32:79–91.
29. Khazaal Y, Chatton A, Bouvard A, *et al*. Internet poker websites and pathological gambling prevention policy. *J Gambl Stud* 2013;29:51–9.
30. Auer M, Malischnig D, Griffiths M. Is "pop-up" messaging in online slot machine gambling effective as a responsible gambling strategy? *Journal of Gambling Issues* 2014;29:1–10.
31. Caillon J, Grall-Bronnec M, Hardouin JB, *et al*. Online gambling's moderators: how effective? Study protocol for a randomized controlled trial. *BMC Public Health* 2015;15:519.

3.4.3 Résumé du protocole

L'objectif du projet EDEIN est de développer un modèle de dépistage des pratiques excessives de jeu en ligne à partir des données de comportements de jeu et d'une évaluation clinique des pratiques de jeu. Le modèle pourra être implémenté dans un outil permettant une analyse en continu des comportements de jeu, afin de repérer les joueurs les plus à risque. Il pourra aussi permettre de mettre en place des mesures de prévention adaptées au profil des joueurs.
Il s'agit d'une étude en trois phases.
La phase 1 consiste à définir une typologie des comportements de jeu en ligne et de repérer les comportements caractéristiques des joueurs potentiellement à risque de développer des problèmes de jeu ; il est prévu l'inclusion de 10 000 joueurs sélectionnés au hasard parmi les clients des opérateurs agréés par l'ARJEL et de 10 000 joueurs parmi les clients de la FDJ.
La phase 2 concerne le développement et l'évaluation d'un score de dépistage à partir des données d'activité de jeu et du score de l'ICJE ; deux autres échantillons (ARJEL et FDJ) de 10 000 joueurs sont prévus. Dans cette étape, les données de l'ARJEL et de la FDJ seront agrégées grâce un identifiant de joueur unique. Ainsi, l'ensemble de l'activité de jeu en ligne légale en France des participants sera prise en compte.
La phase 3 consiste à valider cliniquement le score grâce à des entretiens téléphoniques incluant un diagnostic des troubles liés au jeu ; il est prévu d'inclure 240 joueurs pour cette phase. Pour l'ensemble des trois étapes, les données d'activité de jeu concerneront une période de douze mois, identique pour tous les joueurs ; ainsi nous pourrons analyser les données de joueurs inscrits récemment (pendant la période de douze mois) et des données de joueurs plus anciens (inscrits avant la période de douze mois). De plus, toutes les analyses seront effectuées en tenant compte des différents types de jeu (paris sportifs, paris hippiques, poker, jeux de tirage et de grattage).

3.4.4 Difficultés liées à l'avancement du projet

Comme introduit au début du présent manuscrit, les objectifs de la thèse sont en lien avec l'étude EDEIN, c'est pourquoi nous présentons ici l'état d'avancement de l'étude et ses conséquences sur une partie des résultats de la thèse.
En effet, les résultats présentés dans le chapitre 5 ne répondent pas en totalité aux objectifs définis dans le projet EDEIN puisqu'ils sont basés uniquement sur l'analyse des données d'activité de jeu fournies par la FDJ dans le cadre de la phase 1 en février 2017, seules données à notre disposition. Nous n'avons effectivement pas eu accès aux données des autres opérateurs de jeu par l'intermédiaire de l'ARJEL, ni aux données du questionnaire ICJE. Concernant la phase 3, le recrutement des joueurs a néanmoins pu être lancé. La principale difficulté concerne les demandes d'autorisations déposées auprès de la Commission nationale de l'informatique et des libertés (CNIL) et nécessaires à la transmission et au traitement des données d'activité de jeu. En mars 2018, suite à une requalification du dossier déposé par le CHU de Nantes, la CNIL a autorisé la mise en œuvre du traitement des données. L'ARJEL n'a à ce jour pas eu de réponse

à sa demande d'autorisation CNIL devant lui permettre d'ajouter le CHU de Nantes en tant que destinataire des données. De son côté, la FDJ doit mettre en place une procédure permettant de faire le lien entre les réponses à l'ICJE (récoltées par le CHU de Nantes) et les données d'activité de jeu (stockées par la FDJ) des joueurs ayant répondu au questionnaire. Des échanges entre le CHU de Nantes, l'ARJEL, la FDJ et la CNIL sont encore en cours pour trouver des solutions permettant de mener à bien le projet.
En conséquence, dans la suite du manuscrit, nous décrivons uniquement les méthodes et les résultats relatifs à la phase 1, basés sur les données d'activité des jeux de tirage et de grattage des joueurs la FDJ.

3.4.5 Structure des données de la FDJ

Les données d'activité de jeu à notre disposition proviennent d'un échantillon de 10 000 joueurs possédant un compte chez la FDJ. Les données ont été collectées rétrospectivement par la FDJ, par tirage aléatoire des joueurs sur la période septembre 2015-août 2016. Pour être inclus, les joueurs devaient avoir joué à au moins un jeu de tirage ou de grattage durant cette période. Le résultats de l'extraction nous a été transmis par la FDJ sous forme de plusieurs fichiers contenant des informations sur :

les joueurs (une ligne par joueur, 10 000 lignes) : identifiant du joueur, date d'inscription, âge à la date de collecte des données, sexe ;

les prises de jeu (une ligne par joueur et par nom de jeu, 483 074 lignes) : identifiant du joueur, date de la prise de jeu, type de jeu (grattage, tirage différé, tirage immédiat), nom du jeu, nombre de prises de jeu (pour ce jeu), nombre de gains, montant des gains, montant des mises, montant des pertes, nombre de tirages différés effectués dans le cadre d'un abonnement ;

les versements effectués sur le compte (une ligne par joueur et par date de versement, 100 478 lignes) : identifiant du joueur, date du versement, nombre de versements, montant des versements ;

les retraits (une ligne par client et par date de retrait, 11 lignes) : identifiant du joueur, date du retrait, nombre de retraits, montant des retraits ;

le premier proxy de *chasing* (une ligne par joueur et pour chaque semaine pendant laquelle au moins un versement a été effectué, 80 949 lignes) : identifiant du joueur, numéro de semaine, nombre de versements, montant des versements, nombre d'événements "3 versements effectués dans une période glissante de 12 heures" ;

le second proxy de *chasing* (une ligne par joueur et par semaine, 174 988 lignes) : identifiant du joueur, numéro de semaine, nombre d'événements "versement effectué dans l'heure qui suit une prise de jeu de type hasard pur", nombre d'événements "versement effectué dans l'heure qui suit une prise de jeu de type hasard pur ou de type pari sportif" ;

les bonus de bienvenue et de fidélité (une ligne par joueur et par date de versement du bonus, 10 928 lignes) : identifiant du joueur, date de versement du bonus, nombre de bonus versés, montant des bonus versés, montant des bonus utilisés, nombre de bonus utilisés ;

auto-exclusion (une ligne par joueur auto-exclu, 312 lignes) : identifiant du joueur, date de début de l'auto-exclusion, date de fin de l'auto-exclusion.

Il convient de noter que, si les données relatives aux prises de jeu concernent ici uniquement les jeux de tirage et de grattage, les données liées aux versements et au premier proxy de *chasing* sont indépendantes du type de jeu (paris sportifs ou jeux de hasard pur), puisqu'on ne peut pas savoir à quel type de jeu est destiné un versement d'argent.

3.4.6 Choix méthodologiques et ajustements par rapport au *study protocol*

Comme expliqué dans le *study protocol* de l'étude EDEIN, nous avons défini comme unité statistique l'activité de jeu mensuelle de chaque joueur, que nous appelons "personne-mois". Autrement dit, pour chaque joueur, on observe 12 répétitions (une par mois) de chaque variable. On notera que les joueurs ayant créé leur compte pendant la période d'étude auront moins de 12 répétitions. L'objectif consistant à réaliser une typologie des comportements de jeu revient donc à établir une classification des "personnes-mois". Par rapport à ce qui est décrit dans le protocole, nous avons fait le choix de considérer, en plus, le niveau "joueur" et ainsi travailler dans un contexte de données multi-niveaux (ou imbriquées). Le tableau 3.1 illustre la structure finale des données, après les étapes de data-management. Ainsi, par exemple, le joueur 1 n'a pas versé d'argent sur son compte ni misé d'argent en septembre et en octobre, puis a versé un total de 20 euros et misé pour un total de 15 euros en novembre, etc. Le joueur 2 a créé son compte en février 2016, on ne dispose donc de son activité de jeu uniquement à partir de cette date, jusqu'au mois d'août 2016. Chaque ligne du tableau correspond à un "joueur-mois" (niveau 1) et chaque ensemble de lignes avec le même identifiant de joueur correspond à l'activité d'un joueur (niveau 2). Grâce aux méthodes présentées dans le chapitre 4, nous verrons qu'il est possible de classer simultanément, à partir d'un même modèle générant deux niveaux de classification, les comportements de jeu mensuels (joueurs-mois) et les joueurs eux-mêmes.

Bien que cela ne figure pas dans le *study protocol*, nous nous intéresserons dans la suite du manuscrit au parcours des joueurs dans les mois suivant la création de leur compte. L'objectif de cette analyse, qui concerne donc les joueurs ayant créé leur compte entre septembre 2015 et août 2016, est d'identifier des groupes de joueurs partageant des trajectoires d'activité de jeu similaires. *In fine*, cela pourrait permettre de repérer de manière précoce les joueurs à risque de développer des problèmes de jeu.

En résumé, dans les deux chapitres suivants, nous présentons les méthodes et les résultats liés à deux analyses basées sur les données de comportement de jeu des clients de la FDJ : 1) une classification des comportements de jeu ; 2) une analyse des trajectoires d'activité de jeu dans les six mois suivant l'ouverture d'un compte-joueur.

TABLE 3.1 – Illustration de la structure des données d'activité de jeu

ID Joueur	Mois	Versements	Argent misé	...
1	Sept	0	0	...
1	Oct	0	0	...
1	Nov	20	15	...
1	Déc	0	2	...
1	Janv	0	0	...
1	Févr	0	0	...
1	Mars	10	8	...
1	Avr	20	10	...
1	Mai	0	0	...
1	Juin	0	0	...
1	Juill	10	12	...
1	Août	0	0	...
2	Févr	10	14	...
2	Mars	32	36	...
2	Avr	30	60	...
2	Mai	35	64	...
2	Juin	0	0	...
2	Juill	30	25	...
2	Août	25	32	...

Chapitre 4

Méthodes statistiques

Sommaire

4.1 Classification basée sur les modèles à classes latentes 48
4.1.1 Les différentes méthodes de classification non-supervisée 48
4.1.2 Classification basée sur les modèles de mélange (ou sur les classes latentes) 49
Cas général de la classification basée sur les modèles de mélange 50
Classification basée sur les mélanges de lois normales 50
Analyse en classes latentes 52
Latent Class Cluster Analysis 54
4.1.3 Estimation des paramètres 55
4.1.4 Sélection de modèle 56
4.1.5 Modèles à classes latentes multi-niveaux 58
Principe 58
Écriture et estimation du modèle 59
4.2 Modèles mixtes à classes latentes 61
4.2.1 Principe 61
4.2.2 Écriture des modèles 61
Modèle linéaire mixte (*growth model*) 61
Modèle de trajectoires basé sur les groupes (*group-based trajectory model*) 61
Modèle mixte à classes latentes (*growth mixture model*) 62
Estimation du modèle mixte à classes latentes 63

Ce chapitre présente dans un premier temps le modèle de classification à classes latentes multi-niveaux utilisé pour établir une typologie des comportements de jeu en ligne. Dans un second temps, nous introduisons les modèles de trajectoires basées sur les classes latentes qui seront utilisés dans la section 5.2 pour modéliser l'évolution des comportements de jeu des nouveaux joueurs.

4.1 Classification basée sur les modèles à classes latentes

4.1.1 Les différentes méthodes de classification non-supervisée

De manière générale, la classification (*clustering*) peut être définie comme une procédure consistant à diviser des données en un petit groupe d'observations dont les caractéristiques permettent de décrire de manière concise les similitudes et les différences présentes dans les données (Everitt, Landau, Leese, & Stahl, 2011). Le nombre de classes et leurs propriétés sont inconnues *a piori*. Afin de lever une possible ambiguïté, on peut préciser qu'on parle ici de "classification non supervisée", puisqu'on ne connait pas le groupe réel auquel appartient chaque observation. Ce type de classification est désignée par le terme "*clustering*" en Anglais, alors que le terme "*classification*" (en Anglais) désigne généralement la classification dite supervisée. La classification supervisée consiste à s'appuyer sur des observations dont on connait le groupe réel, pour construire un modèle qui servira dans la plupart des cas à classer de nouvelles observations dont on ne connait pas le groupe. Dans la suite du manuscrit, nous emploierons le terme "classification" pour désigner la "classification non supervisée". De plus, par souci de cohérence avec les définitions données à partir de la section 4.1.5, l'anglicisme "cluster" sera utilisé pour désigner les groupes d'observations (le mot "groupe" pourra aussi être utilisé), mais on évitera de parler de "classes" en dehors de la dénomination des méthodes (jusqu'à la section 4.1.5).
Il existe de nombreuses techniques pour réaliser une classification des observations. L'ouvrage de Everitt et al. (2011) propose une revue d'une partie de ces méthodes. Les plus utilisées sont la Classification Ascendante Hiérarchique (CAH), l'algorithme des *k-means*, et les méthodes basées sur les modèles de mélange. Chacune de ces méthodes possède des variantes qui sont notamment listées par Leisch et Gruen (2018). Nous décrivons ici succinctement le principe de la CAH et de l'algorithme des *k-means* La classification basée sur les modèles de mélange fait l'objet des sections suivantes.
La CAH consiste en une série de partitionnements successifs des observations. Au début de l'algorithme, toutes les observations sont dans un cluster différent ; si l'algorithme est exécuté en entier, à la dernière itération, toutes les observations appartiennent à un unique et même cluster. A chaque itération, l'algorithme fusionne les observations qui sont les plus "proches" (selon une mesure de distance définie en amont) ; les clusters ainsi formés à chaque itération sont alors successivement fusionnés au fur et à mesure de l'exécution de l'algorithme. La définition de la distance entre deux clusters fait aussi l'objet d'un choix à faire en amont de l'analyse. Une re-

présentation graphique de la procédure complète de classification à l'aide d'un dendrogramme permet de déterminer le nombre de clusters qui parait optimal.
L'algorithme des *k-means* fonctionne de la manière suivante : 1) on définit *a priori* un nombre de clusters K puis on sélectionne aléatoirement K observations qui correspondent aux "centres initiaux" ; 2) on calcule la distance entre chaque observation et chaque centre (on se base généralement sur la distance euclidienne, mais d'autres distances peuvent être utilisées) ; 3) chaque observation est affectée au centre le plus proche ; 4) on obtient alors des groupes dont on calcule le "centre de gravité", c'est à dire l'observation fictive qui possède les caractéristiques moyennes du groupe. Un nouveau centre est ainsi défini pour chaque groupe ; 5) on répète les étapes 1 à 4 jusqu'à ce que les observations ne changent plus de groupe entre deux itérations.
Les méthodes de classification hiérarchiques (CAH) ou basées sur les centroïdes (*k-means*) sont des procédures heuristiques globalement utiles mais qui présentent plusieurs inconvénients. Tout d'abord, plusieurs choix sont à faire quant au nombre de clusters et à la façon de calculer la distance entre les observations. Or, il n'existe pas de lignes directrices claires indiquant les meilleures options. Ensuite, les deux méthodes ont été développées dans le cadre de variables quantitatives continues et leur application à des variables discrètes nécessite de transformer les données, typiquement à l'aide d'une analyse des correspondances multiples. De plus, toutes les variables doivent être mesurées sur la même échelle puisque des variables mesurées dans des unités différentes invalideraient les calculs de mesure des distances. Les données doivent donc être standardisées avant l'analyse. Un autre inconvénient propre aux *k-means* est la sélection aléatoire des centres initiaux, qui peut entraîner des partitions finales différentes pour un même ensemble de données. Par ailleurs, dans les deux méthodes, chaque observation est affectée à un cluster de manière déterministe, sans prendre en compte les probabilités d'appartenance aux autres clusters. Enfin, il n'existe pas de critère clair permettant d'évaluer la "qualité" de la classification.
Une méthode alternative est la classification basée sur l'estimation d'un modèle de mélange, dont la nature probabiliste permet de répondre à plusieurs problèmes des méthodes de classification plus classiques.

4.1.2 Classification basée sur les modèles de mélange (ou sur les classes latentes)

Nous introduisons ici les méthodes de classification basées sur les modèles de mélange. Nous présentons d'abord le cas général (indépendamment de la distribution des variables), puis le cas des mélanges de lois normales, et enfin le cas des mélanges de lois binomiales et multinomiales (analyse en classes latentes). Ce déroulement permettra de présenter plus naturellement la classification basée sur le mélange de lois issues de familles paramétriques différentes (section 4.1.2, dont l'extension à des données multi-niveaux (section 4.1.5) constitue le modèle final de classification utilisé dans cette thèse.

Cas général de la classification basée sur les modèles de mélange

La classification basée sur les modèles de mélange repose sur l'hypothèse que les observations de l'échantillon proviennent d'une population composée de plusieurs sous-populations (clusters). Chaque cluster est défini par une densité de probabilité multivariée ; le mélange de ces densités définit la densité de probabilité qui décrit la population. L'estimation des paramètres du modèle de mélange permet de calculer la probabilité que chaque observation a d'appartenir aux différents clusters. Par rapport aux algorithmes de classification présentées précédemment, cette méthode est basée sur un modèle statistique dont on peut mesurer les propriétés. Cela permet de comparer plusieurs modèles à partir de critères statistiques, pour notamment sélectionner le nombre de clusters et évaluer la qualité de la classification. De plus, il n'est pas nécessaire de standardiser les variables et le modèle permet de s'affranchir des problèmes liés à la mesure de distances entre les observations.
Il existe plusieurs termes pour désigner cette méthode et ses variantes, en fonction de la nature des variables et des domaines de recherche. Le terme Anglais *model-based clustering* (Fraley & Raftery, 1998, 2002) définit le modèle le plus "général" puisqu'il désigne un modèle de mélange fini que l'on peut écrire :

$$f(y_i|\theta) = \sum_{k=1}^{K} \pi_k f_k(y_i|\theta_k) \tag{4.1}$$

où y_i désigne, pour l'observation i, le vecteur des valeurs observées pour l'ensemble des variables, K est le nombre de clusters, π_k est la probabilité a priori d'appartenir au cluster k (autrement dit, π_k correspond à la "taille" du cluster k), θ désigne l'ensemble des paramètres du modèle de mélange, et θ_k correspond aux paramètres de la fonction de densité $f_k(y_i|\theta_k)$ spécifique au cluster k. Appliqué à un objectif de classification, ce modèle est aussi appelé, plus rarement, *mixture-model clustering*.

Classification basée sur les mélanges de lois normales

Un cas particulier des modèles de mélange qui a été largement étudié est celui des mélanges de lois gaussiennes. On utilise alors le terme *gaussian mixture model*, ou *model-based gaussian clustering* dans un contexte de classification (Banfield & Raftery, 1993). Le modèle s'écrit :

$$f(y_i|\theta) = \sum_{k=1}^{K} \pi_k \phi_k(y_i|\mu_k, \Sigma_k) \tag{4.2}$$

où ϕ_k désigne la densité de probabilité d'une loi normale multivariée de paramètres μ_k et Σ_k. On notera que lorsque le nombre de variables ou de clusters augmente, le nombre d'éléments de la matrice de variance-covariance Σ_k croit rapidement et peut poser des problèmes lors de l'estima-

tion du modèle. Des solutions ont alors été envisagées pour réduire le nombre de paramètres à estimer, en simplifiant la forme de Σ_k. Banfield et Raftery (1993) puis Celeux et Govaert (1995) ont proposé une décomposition en valeurs propres de Σ_k permettant plusieurs paramétrisations de la matrice de variance-covariance. Cela permet de définir plusieurs modèles allant du plus simple où tous les clusters partagent les mêmes caractéristiques, au plus complexe où tous les clusters peuvent être différents. La décomposition proposée est la suivante :

$$\Sigma_k = \lambda_k D_k A_k D_k' \tag{4.3}$$

où λ_k est une constante, D_k est la matrice des vecteurs propres de Σ_k, D_k est sa transposée, et A_k est une matrice diagonale dont les valeurs sont proportionnelles aux valeurs propres de Σ_k. On peut donner une interprétation géométrique aux éléments issues de la décomposition : λ_k correspond au volume (taille) du cluster k, D_k contrôle l'orientation du cluster et A_k définit sa forme. La figure 4.1 réalisée avec le package R `mclust` (Scrucca, Fop, Murphy, & Raftery, 2016) illustre ces propriétés géométriques en prenant comme exemple une classification effectuée à partir des données des iris de Fisher (1936). La figure 4.1 représente graphiquement trois clusters dans un espace à deux dimensions défini par les variables "Sepal.Length" et "Petal.Length". Dans le graphique (1), aucune contrainte n'est posée sur la structure de Σ_k, les trois clusters ont un donc un volume différent, une forme différente, et une orientation différente. Dans le graphique (2), le volume des clusters et laissée libre, mais leur forme et leur orientation sont identiques. Dans le graphique (3), les trois clusters ont le même volume, leur forme est contrainte à être sphérique et ils n'ont donc pas d'orientation. Fraley et Raftery (2003) et Scrucca et al. (2016) ont étendu les travaux de Banfield et Raftery puis Celeux et Govaert pour proposer au total quatorze paramétrisations de la matrice de variance-covariance Σ_k, implémentées dans MCLUST. Pour n'évoquer que celles illustrées dans la figure 4.1, la structure de Σ_k la plus complexe, notée "VVV" pour "volume Variable, forme Variable, orientation Variable" s'écrit $\lambda_k D_k A_k D_k'$, autrement dit, aucune contrainte n'est ajoutée par rapport à l'équation (4.3). La structure "VEE" pour "volume Variable, forme Egale, orientation Egale" s'écrit $\lambda_k DAD'$. La structure la plus simple pour Σ_k, "EII", s'écrit λI et impose que les clusters soient de la même taille et de forme sphérique. Il est intéressant de noter que cette dernière paramétrisation fait des hypothèses très proches de celles faites par l'algorithme des *k-means*, et que, malgré des procédures d'estimation différentes, les deux méthodes sont quasiment équivalentes (Fraley & Raftery, 2002).
Une spécification intéressante est celle où $\Sigma_k = \lambda_k A_k$, c'est à dire où Σ_k est une matrice diagonale dans chaque cluster. Dans cette paramétrisation, on suppose que les covariances entre les variables dans un même cluster sont égales à zéro ; c'est l'hypothèse d'indépendance locale. Les variances sont quant à elles estimées dans chaque cluster, ou bien considérées égales dans tous les clusters (dans ce dernier cas, $\Sigma_k = \Sigma$).

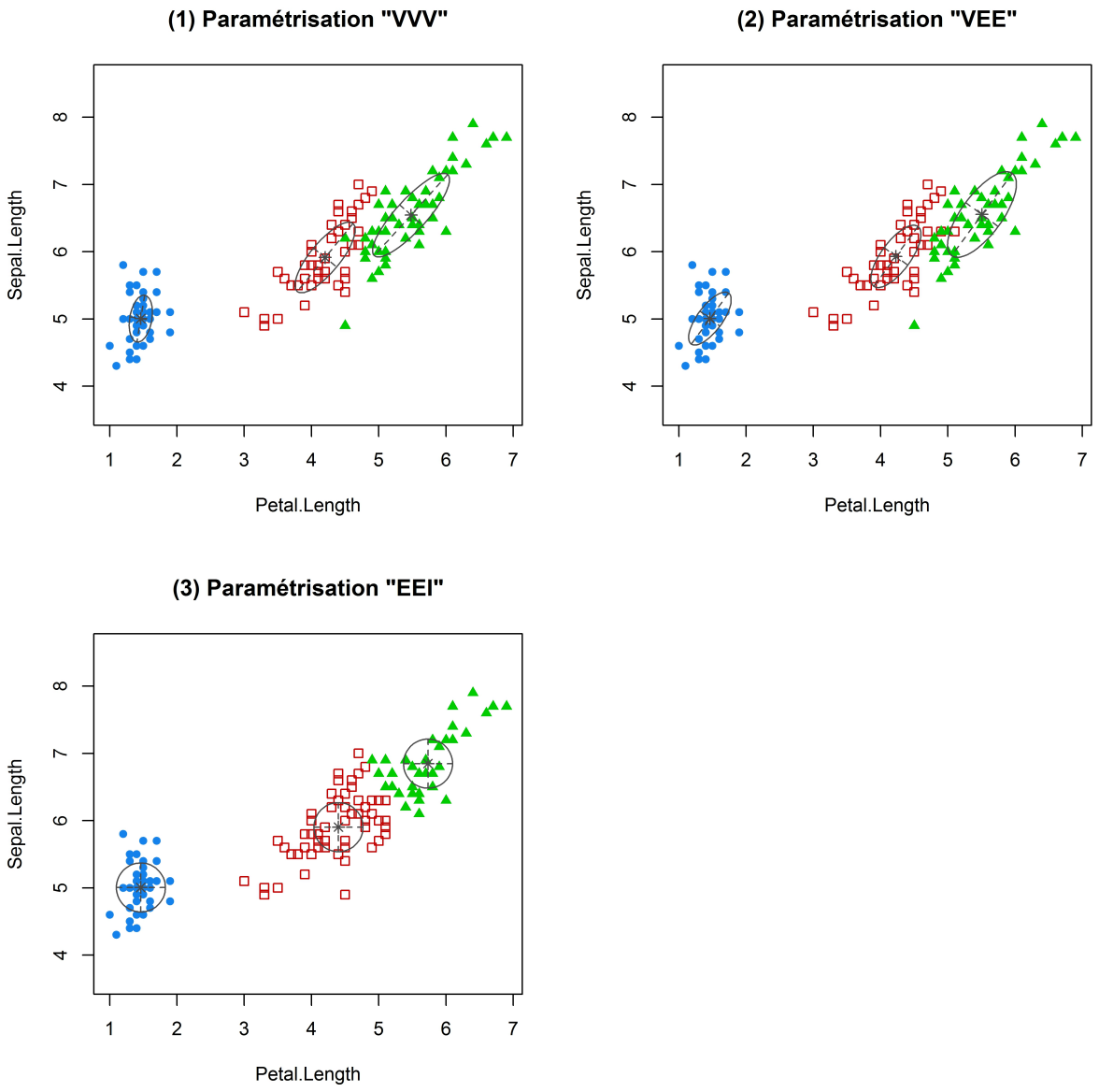


FIGURE 4.1 – Exemple de représentation graphique de clusters en fonction de la paramétrisation de la matrice de variance-covariance dans un mélange de lois normales

Analyse en classes latentes

L'Analyse en Classes Latentes (*Latent Class Analysis*) (LCA), au sens de la définition de (McCutcheon, 1987), consiste à analyser les réponses à des variables catégorielles dont les relations sont expliquées par l'appartenance des répondants à des classes latentes mutuellement exclusives. Le terme "latent" signifie que les groupes (classes) sont *a priori* inconnues. Dans un contexte de classification, l'objectif de la LCA est d'identifier ces groupes et d'estimer la proba-

TABLE 4.1 – Noms de différents types de modèles à variable latente

		Variable latente	
		Continue	**Discrète**
Variables observées	**Continues**	Factor analysis	Latent profile analysis Model-based clustering Mixture model clustering Latent class cluster analysis
	Discrètes	Latent trait model Item response theory	Latent class analysis

bilité que chaque individu a d'appartenir à ces groupes.
Initialement développée dans les années 1950 par Lazarsfeld pour analyser des données dichotomiques, la méthode a été étendue à l'analyse de variables polytomiques, grâce notamment à L. A. Goodman (1974) qui a formalisé qui a généralisé le problème d'analyse des tableaux de contingence en un problème d'estimation de paramètres d'un modèle probabiliste. Le lien entre la LCA et les méthodes de classification basées sur les modèles de mélanges présentées précédemment a été mis en évidence par (Vermunt & Magidson, 2002). En partant de l'équation (4.1), il suffit en effet de spécifier que $f_k(y_i|\theta_k)$ est non plus la densité d'une loi normale multivariée comme dans le cas des mélanges gaussiens (cf. équation (4.2)), mais une loi de probabilité discrète univariée (loi binomiale dans le cas de variables dichotomiques et loi multinomiale pour des données polytomiques) pour chaque variable. L'introduction de l'hypothèse d'indépendance locale est par contre nécessaire, contrairement au cas des données normales où la covariance peut être naturellement modélisée grâce à Σ_k. Le modèle de la LCA pour T variables discrètes peut alors s'écrire :

$$P(y_{i1} = m_1, y_{i2} = m_2, ..., y_{iT} = m_T) = \sum_{k=1}^{K} P(x) \prod_{t=1}^{J} P(y_{it} = m_t|x) \tag{4.4}$$

où x désigne l'appartenance à un cluster.
La différence de dénomination entre les deux méthodes (*model-based clustering* et LCA) réside finalement dans la nature (continue ou discrète) des variables observées. On notera que le terme "*latent profile analysis*" est aussi parfois utilisé pour désigner le cas où les variables observées sont continues. Pour aller un peu plus loin, il est aussi possible de nommer différents modèles en fonction de la nature de la variable latente. Le tableau 4.1 présente, de manière non exhaustive, les noms (en Anglais) de ces modèles. Dans notre cas, étant donné l'objectif de classification, nous nous intéressons à la partie droite du tableau 4.1, c'est à dire les modèles à variable latente

discrète. Nous allons voir dans la section suivante qu'on peut définir un modèle à variable latente discrète plus général, qui peut s'appliquer à un ensemble de variables discrètes et continues.

Latent Class Cluster Analysis

Vermunt et Magidson (2002) ont proposé un modèle à variable latente discrète plus général, qui permet de ne plus avoir à distinguer les modèles de classification selon la nature des variables observées, et permet surtout d'estimer un modèle de classification à partir de variables de différentes natures. La méthode est nommée Latent Class Cluster Analysis (LCCA) par ses auteurs, ce que nous pouvons traduire en Français par "classification basée sur classes latentes" (ou "classification à partir de classes latentes")
Sous l'hypothèse d'indépendance locale, le modèle a la forme suivante :

$$f(y_i|\theta) = \sum_{k=1}^{K} \pi_k \prod_{t=1}^{J} f_k(y_{it}|\theta_{tk}) \tag{4.5}$$

où T désigne le nombre de variables et t désigne une variable particulière. L'idée est de spécifier la forme $f_k(y_i t|\theta_t k)$ pour chaque variable j. En pratique on peut par exemple spécifier une loi normale pour une variable continue, une loi binomiale pour une variable dichotomique, une loi multinomiale pour une variable polytomique nominale, et une loi de Poisson pour une variable mesurant un nombre d'événements dans un intervalle de temps. Il est important de noter qu'il s'agit de distributions cluster-spécifiques, c'est à dire que, dans le cas, par exemple, d'une variable continue, on fait l'hypothèse que la variable est distribuée normalement *dans chaque cluster*, mais que sa distribution marginale est un mélange de loi normales univariées.
Si on considère le cas où toutes les variables sont continues et supposées normalement distribuées dans les clusters (comme dans la section 4.1.2), le modèle défini par l'équation (4.5) peut être rendu équivalent au modèle (4.2) (c'est à dire un modèle avec une seule loi normale multivariée et sans l'hypothèse d'indépendance locale) grâce à une légère modification. Il suffit en effet de ne plus considérer l'indice t comme désignant une variable unique mais comme un indice représentant un ensemble de variables. De cette manière, toutes les variables continues peuvent être désignées par un t particulier et être modélisées par une loi normale multivariée (où les covariances intra-clusters peuvent être estimées, c'est à dire que l'hypothèse d'indépendance locale n'est pas nécessaire). Pour les variables discrètes, chacune peut être désignée par un indice t différent et on spécifie alors autant de lois de probabilité discrètes univariées, en faisant l'hypothèse d'indépendance locale. Il est également possible de relâcher la contrainte d'indépendance locale pour des variables discrètes en spécifiant une distribution multinomiale multivariée ou pour des nombres d'événements en spécifiant une loi de Poisson multivariée. Dans le cas d'une loi normale multivariée, il est possible de considérer des covariances intra-clusters uniquement entre certaines variables en spécifiant des éléments non nuls en dehors de la diagonale de Σ_k.

4.1.3 Estimation des paramètres

L'estimation des paramètres d'un modèle d'un mélange se fait généralement par maximisation de la log-vraisemblance suivante :

$$\log(\mathcal{L}(\Theta|\mathcal{Y})) = \sum_{i=1}^{N}\sum_{k=1}^{K}\log(\pi_k f_k(y_i|\theta_k)) \tag{4.6}$$

où Θ désigne l'ensemble des paramètres à estimer et $\mathcal{Y}$ l'ensemble des données. Cette vraisemblance est difficile à maximiser à l'aide d'un algorithme du type Newton-Raphson. Une solution est alors d'utiliser l'algorithme Expectation-Maximization (EM) (Dempster, Laird, & Rubin, 1977) qui permet de considérer la log-vraisemblance "complète", c'est à dire en supposant que l'on connait l'appartenance des observations aux clusters. Si on note $\mathcal{X} = (x_i)_{i=1,\dots,N}$ et qu'on suppose que $x_i = k$ si l'observation i appartient au cluster k, alors la log-vraisemblance complète s'écrit :

$$\log(\mathcal{L}(\Theta|\mathcal{Y},\mathcal{X})) = \sum_{i=1}^{N}\sum_{k=1}^{K}x_{ik}\log(\pi_k f(y_i|\theta)) = \sum_{i=1}^{N}\sum_{k=1}^{K}x_{ik}(log\pi_k + \log(f(y_i|\theta))) \tag{4.7}$$

A chaque étape E d'un cycle g de l'algorithme, les estimations $\hat{x}_{ik}$ de x_{ik} sont obtenues en remplaçant Θ par son estimation $\hat{\Theta}^{(g-1)}$ calculée lors du cycle précédent. On peut estimer $x_{ik}^{(g)}$ grâce au théorème de Bayes : $\pi_{ik}^{(g)} = \hat{x}_{ik}^{(g)} = \frac{\pi_k^{(g)} f_k(y_i|\theta_k^{(g)})}{f(y_i|\Theta^{(g)})} = \frac{\pi_k^{(g)} f_k(y_i|\theta_k^{(g)})}{\sum_{l=1}^{K}\pi_l^{(g)} f_k(y_i|\theta_l^{(g)})}$. A l'étape M, on cherche un nouveau $\hat{\Theta}^g$ qui maximise $log(\mathcal{L}(\Theta|\mathcal{Y},\mathcal{X}))$. Au final, on montre que $\hat{\pi}_l^{(g)} = \frac{1}{N}\sum_{i=1}^{N}\pi_{li}^{(g-1)}$. L'estimation des paramètres des densités de probabilité spécifiques aux clusters-dépend de la nature de la loi spécifiée. Dans la cas gaussien, on montre que $\mu_l^{(g)} = \frac{\sum_{i=1}^{N} y_i \pi_{li}^{(g-1)}}{\sum_{i=1}^{N}\pi_{li}^{(g-1)}}$ et $\Sigma_l^{(g)} = \frac{\sum_{i=1}^{N}\pi_{li}^{(g-1)}(y_i-\mu_l^{(g-1)})(y_i-\mu_l^{(g-1)})'}{\sum_{i=1}^{N}\pi_{li}^{(g-1)}}$. Dans le cas d'une loi binomiale (LCA, modèle (4.4)), $\hat{p}_{jl} = \frac{\sum_{i=1}^{N} y_{ij}\pi_{li}^{(g-1)}}{\sum_{i=1}^{N}\pi_{li}^{(g-1)}}$.

Vermunt et Magidson (2015) ont proposé une méthode de maximisation de la vraisemblance qui combine l'algorithme EM et l'algorithme Newton-Raphson. Cela permet de tirer profit de la stabilité du premier algorithme au début de l'optimisation et de la vitesse de convergence du second lorsqu'on se rapproche d'une solution optimale. Par ailleurs, il a aussi été proposé une méthode d'estimation bayésienne alternative au maximum de vraisemblance, le "maximum a posteriori", (Maris, 1999 ; Garre & Vermunt, 2006). L'idée est de spécifier des lois a priori pour les paramètres et de maximiser une fonction qui est la somme de la fonction de log-vraisemblance

et des logarithmes des lois a priori. Les informations des loi a priori permettent d'éviter d'estimer des valeurs de paramètres situés à la frontière de l'espace des paramètres, où la fonction de vraisemblance peut atteindre un maximum (c'est à dire, par exemple, un espace qui correspondrait à des probabilités de réponse strictement égales à 0 ou 1 pour des variables binaires, ou des tailles de clusters égales à 0 ou 1). En utilisant des lois a priori "non informatives", la méthode permet de retrouver des résultats équivalents à ceux obtenus avec la méthode du maximum de vraisemblance si celle-ci évite les points de singularité (Garre & Vermunt, 2006).
Par ailleurs, l'estimation des paramètres d'un modèle de mélange est connue pour être associée à des problèmes de maxima locaux. Il est donc conseillé de relancer la procédure d'optimisation plusieurs fois, avec des valeurs initiales différentes pour $\theta^{(0)}$.

4.1.4 Sélection de modèle

La sélection d'un modèle de classification basé sur les classes latentes implique deux problèmes principaux : le choix du nombre de clusters (que l'on considérait jusqu'ici comme fixé) et la "forme" des clusters. Le second problème est plutôt caractéristique du cas des mélanges gaussiens, et consiste à trouver la "meilleure" structure de la matrice de variance-covariance Σ_k. Nous avons décrit dans la section 4.1.2 différentes manières de spécifier Σ_k. Pour un nombre de clusters fixé, il est possible de comparer deux modèles avec une structure différente de Σ_k grâce, par exemple, à un test de rapport de vraisemblance.
Concernant le choix du nombre de clusters, deux modèles avec un nombre de clusters différents, bien que pouvant être considérés comme emboités, ne peuvent être comparés à l'aide d'un test de rapport de vraisemblance parce que la distribution de la statistique de test de suit pas asymptotiquement une loi du χ^2 (puisque sous l'hypothèse nulle, certains des paramètres π_k désignant la taille des clusters seraient supposés être égaux à 0 (ou 1), et le modèle ne serait pas donc identifiable). Le choix du nombre de clusters se fait donc plutôt grâce à des critères d'information, les plus utilisés étant le Akaike Information Criterion (AIC) : $AIC = -2log\mathcal{L} + 2npar$ et le Bayesian Information Criterion (BIC) : $BIC = -2log\mathcal{L} + log(N)npar$. Ces critères sont des fonctions de la log-vraisemblance et du nombre de paramètres *npar* à estimer dans le modèle, et sont donc des indices d'adéquation du modèle qui doivent être comparer entre eux. D'autres critères permettent de mesurer la "performance" ou la "précision" des classifications *a posteriori*. Si on écrit les classifications a posteriori de cette façon : $\hat{P}(x_i|y_i) = \frac{\hat{P}(x_i)\hat{f}(y_i|x)}{\hat{f}(y_i)}$. On peut alors par exemple calculer le "taux d'erreur de classifications" : $Err = \frac{\sum_{i=1}^{N}(1-\max\hat{P}(x_i|y_i))}{N}$. L'indice est en théorie compris entre 0 et 1. Plus il est proche de 0, plus la "certidude" des classifications individuelles est grande. Par exemple, s'il vaut 0,1, cela signifie que chaque observation a en moyenne 90% de chance d'appartenir à un cluster particulier (et un pourcentage cumulé de 10% de chance d'être dans les autres clusters). Un deuxième indice similaire, l'entropie (Celeux & Soromenho, 1996), mesure également les performances de classification par la formule suivante : $Entropie = \sum_{i=1}^{N}\sum_{x=1}^{K} -\hat{P}(x_i|y_i)log\hat{P}(x_i|y_i)$. Plus l'entropie est proche de 1, plus les clusters

sont bien "délimités". L'entropie permet de définir un critère indiquant à la fois l'adéquation du modèle et la qualité de la classification : le critère Approximate Weight of Evidence (AWE) (Banfield & Raftery, 1993), dont la formule est : $AWE = -2log\mathcal{L}+2Entropie+2(\frac{3}{2}+logN)npar$. Pour l'AIC, le BIC et l'AWE, des valeurs faibles indiquent un meilleur modèle.
Outre les critères d'information et les indices de la qualité de la classification, des tests statistiques permettent de comparer un modèle composé $k+1$ clusters à un modèle composé de k clusters. Lo, Mendell, et Rubin (2001) ont ainsi développé le test Lo-Mendel-Rubin Likelihood-Ratio Test (LMR), basé une distribution corrigée de la statistique de test $-2log(\mathcal{L}_{(K)} - \mathcal{L}_{(K+1)})$ (la fonction de distribution est décrite dans l'article de Lo et al.). Une p-value faible (ex : <0,05) rejette l'hypothèse que le modèle à K clusters est meilleur que le modèle à $K+1$ clusters. Un autre test, le Bootstrap Likelihood-Ratio Test (BLRT) (McLachlan & Peel, 2000), permet de comparer deux modèles grâce un échantillonnage par bootstrap. Le principe du BLRT est d'estimer les paramètres $\Theta_{(K)}$ d'un modèle à K clusters, et de générer un échantillon à partir de $\hat{\Theta}_(K)$ sous l'hypothèse que le modèle à K classes est le modèle adéquat. On analyse ensuite cet échantillon avec un modèle à K clusters et un modèle à $K+1$ clusters. La procédure est répétée B fois (ex : 500 répétitions) puis on calcule le rapport des log-vraisemblances pour chaque échantillon. On compte ensuite la proportion de rapports de log-vraisemblances supérieurs au rapport des log-vraisemblances calculé à partir de l'échantillon original (analysé avec un modèle à K clusters et avec un modèle à $K+1$ clusters). Si cette proportion est faible (ex : $< 5\%$), alors on rejette l'hypothèse que le modèle à K clusters est meilleur que le modèle à $K+1$ clusters.
Il existe donc de nombreux critères statistiques (tous ne sont pas évoqués ici) pour déterminer le nombre optimal de clusters. Nylund, Asparouhov, et Muthén (2007) ont comparé plusieurs de ces méthodes dans une étude de simulation et les résultats suggèrent que le BIC et le BLRT sont les plus performants. En pratique, le BIC est sans doute le critère statistique le plus utilisé.

En plus des critères statistiques, d'autres éléments sont à prendre en compte dans la sélection de modèle. Par exemple, certains auteurs utilisent la taille (relative ou "absolue") du plus petit cluster comme un critère de sélection, avançant qu'un modèle contenant un cluster trop petit (ex : $< 1\%$) n'est pas approprié. L'argument semble néanmoins davantage relever de considérations pratiques que théoriques et aucune règle n'est clairement définie à ce sujet.
Un autre critère de décision est la reproductibilité du modèle. La reproductibilité désigne la "robustesse" du modèle à des changements des valeurs initiales des paramètres lors du processus d'estimation, mais fait aussi référence au concept de "validation interne". Dans ce cas, il peut par exemple s'agir de séparer l'échantillon en deux, et de comparer les résultats d'un même modèle appliqué aux deux sous-échantillons. Un modèle qui ne serait pas robuste ne devrait donc pas être sélectionné. Une autre manière d'évaluer un modèle et d'étudier la cohérence des clusters avec des informations extérieures, n'ayant pas servi à la classification. Enfin, le choix du nombre de clusters peut se faire en fonction de l'interprétabilité des clusters et de leur pertinence par

rapport au domaine de recherche et à la problématique posée.
Il apparait en pratique que les deux critères les plus utilisés pour sélectionner le modèle final sont le BIC et l'interprétabilité des clusters.

4.1.5 Modèles à classes latentes multi-niveaux

Principe

Une des hypothèses de la LCCA est que les observations sont indépendantes entre elles. Cette hypothèse n'est pas valide s'il s'agit de données imbriquées (par exemple, des groupes de patients "imbriqués" dans centres hospitaliers, ou des groupes d'élèves "imbriqués" dans des écoles) ou de données répétées dans le temps (où les mêmes variables sont mesurées à plusieurs temps différents). Dans ces deux cas, les données des observations d'un même groupe sont en théorie plus corrélées que les données d'observations appartenant à des groupes différents. On appellera les observations du niveau le plus bas (ex : les élèves) les observations de niveau 1 et les observations du niveau le plus haut (ex : les écoles) les observations (ou unités) de niveau 2. Dans le cas de données répétées, le niveau 2 correspond aux individus et le niveau correspond à leurs mesures répétées. Dans notre cas, comme décrit dans la section 3.4.6, le niveau 2 correspond aux joueurs et le niveau 1 correspond à leurs indicateurs d'activité de jeu mesurée chaque mois (joueurs-mois). La LCCA doit donc prendre en compte le fait que le comportement de jeu est plus similaire entre deux temps de mesures pour un même joueur qu'entre deux joueurs différents.
Comme cela est typiquement fait pour l'analyse de données longitudinales, on va donc introduire des effets aléatoires pour prendre en compte le fait que certains paramètres du modèle peuvent varier selon les observations de niveau 2. Autrement dit, on va considérer que chaque joueur a des caractéristiques propres et non observables directement à partir des données mesurées (qu'on peut donc voir comme les valeurs d'une variable latente), qui vont expliquer les différences entre les mesures d'activité de jeu de joueurs différents. Pour donner un exemple, un joueur avec des revenus financiers élevés (variable non mesurée) jouera potentiellement plus d'argent chaque mois qu'un joueur avec des revenus plus faibles, qui aura normalement tendance à miser moins d'argent chaque mois.
On va ici considérer que la variable latente représentant les effets aléatoires est une variable latente discrète. Dans le cas d'une variable latente continue, on ferait l'hypothèse que les paramètres du modèle varient autour d'un effet fixe selon, typiquement, une loi normale. En considérant des effets aléatoires discrets, on va supposer que les effets aléatoires ne peuvent pas prendre n'importe quelle valeur pour chaque individu, mais que le nombre de valeurs possibles des effets aléatoires va se restreindre à un nombre de groupes d'individus fixé. On appellera ces groupes "classes", et on appellera "clusters" les groupes d'observations d'activité de jeu mensuelles. De plus, on va spécifier un modèle ou l'appartenance aux classes n'aura pas d'effet sur les variables de comportements de jeu, mais uniquement sur les probabilités d'apprtenance aux clusters. Cela revient à supposer que les comportements de jeu mensuels des joueurs d'une même classe sont

corrélées parce que ces joueurs ont tendance à tendance, chaque mois, à être dans le même cluster. Et c'est l'appartenance à un cluster qui explique les observations de comportements de jeu mensuels. Cette paramétrisation correspond à un cas particulier d'une LCCA multi-niveaux proposé par (Vermunt, 2003, 2008).

Écriture et estimation du modèle

Pour les observations de niveau 1, le modèle peut s'écrire sous la forme suivante :

$$P(Y_{ij}|Z_j = g) = \sum_{k=1}^{K} P(X_{ij} = k|Z_j = g) \prod_{t=1}^{T} P(Y_{ijt}|X_{ij} = k) \tag{4.8}$$

où $i = 1, ..., I$ désigne les observations de niveau 1 (les joueurs-mois) , $j = 1, ..., J$ désigne les observations de niveau 2 (les joueurs), Y_{ij} désigne les données d'activité de jeu d'un joueur j pour le mois i, $t = 1, ..., T$ désigne les variables observées (ou, comme précisé pour le modèle (4.5), un ensemble de variables modélisées par une loi multivariée) , k désigne une valeur (un cluster) de la variable latente X_{ij} qui regroupe les observations de niveau 1 dans K clusters, et g désigne une valeur (une classe) de la variable latente Z_j qui regroupe les observations de niveau 2 dans G classes.
Le modèle ressemble donc à celui défini par l'équation (4.5), mais ici, les probabilités d'appartenance des observations de comportement de jeu mensuels aux clusters dépendent de la classe d'appartenance des joueurs.
Au niveau 2, le modèle pour les joueurs s'écrit :

$$P(Y_j) = \sum_{g=1}^{G} P(Z_j = g) \prod_{i=1}^{N_j} P(Y_{ij}|Z_j = g) \tag{4.9}$$

où Y_j désigne toutes les observations d'activité de jeu mensuelles d'un joueur j et N_j désigne le nombre d'observations de niveau 1 pour l'observation de niveau 2 j (i.e. le nombre de mois où l'activité de jeu est mesurée pour le joueur j)
En combinant (4.8) et (4.9), on obtient :

$$P(Y_j) = \sum_{g=1}^{G} \left(\underbrace{P(Z_j = g)}_{\text{(a)}} \prod_{i=1}^{N_j} \left(\sum_{k=1}^{K} \underbrace{P(X_{ij} = k|Z_j = g)}_{\text{(b)}} \prod_{t=1}^{T} \underbrace{P(Y_{ijt}|X_{ij} = k)}_{\text{(c)}} \right) \right) \tag{4.10}$$

La composante (a) de l'équation (4.10) correspond à la probabilité qu'un joueur j appartienne à la classe g, la composante (b) correspond à la probabilité qu'un joueur-mois i appartienne au cluster k, sachant que le joueur appartient à la classe g, et la composante (c) représente la probabilité qu'un joueur-mois i ait le vecteur de données Y_{ijt}, sachant qu'il appartient au cluster k. Ainsi, la composante (c) capture les différences entre les comportements de jeu des joueurs-mois

par l'intermédiaire de clusters, et la composante (b) capture les différences entre les joueurs par l'intermédiaire de classes qui varient quant à la taille relative des clusters de joueurs-mois.
Les paramètres du modèle peuvent être estimés grâce à version modifiée de l'étape E de l'algorithme EM (Vermunt, 2003, 2004 ; Bijmolt, Paas, & Vermunt, 2004). La log-vraisemblance basée sur les données complètes d'un joueur j s'écrit :

$$\log(\mathcal{L}_j) = \sum_{g=1}^{G} P\left(Z_j = g|Y_j\right) \log P\left(Z_j = g\right) + \sum_{g=1}^{G}\sum_{k=1}^{K}\sum_{i=1}^{N_j} P\left(Z_j = g, X_{ij} = k|Y_j\right) \log P\left(X_{ij} = k|Z_j = t\right) + \sum_{k=1}^{K}\sum_{i=1}^{N_j}\sum_{t=1}^{T} P\left(X_{ij} = k|Y_j\right) \log P\left(Y_{ijt}|X_{ij} = k\right) \tag{4.11}$$

L'étape E consiste à calculer les probabilités d'appartenance aux classes et aux clusters, étant donné le vecteur d'observation Y_j, c'est à dire $P(Z_j = g, X_{ij} = k|Y_j)$, que l'on peut aussi écrire $P(Z_j = g|Y_j)P(X_{ij} = k|Y_{ij}, Z_j = g)$ ou $P(Z_j = g|Y_j)P(X_{ij} = k|Y_{ij}, Z_j = g)$ puisqu'on suppose que la probabilité d'appartenance d'un joueur-mois à un cluster ne dépend pas des autres observations mensuelles sachant la classe d'appartenance du joueur. $P(X_{ij} = k|Y_{ij}, Z_j = g)$ peut s'écrire $\frac{P(X_{ij}=k, Y_{ij}|Z_j=g)}{P(Y_{ij}|Z_j=k)}$ où $P(X_{ij} = k, Y_{ij}|Z_j = g) = P(X_{ij} = k|Z_j = g)\prod_{t=1}^{T} P(Y_{ijt}|X_{ij} = k)$ puisqu'on suppose que Y_{ij} dépend de k mais pas de g. Concernant le terme $P(Z_j = g|Y_j)$, on peut l'écrire $\frac{P(Z_j=g, Y_j)}{P(Y_j)}$ ou encore $\frac{P(Y_{ij}|Z_j=g)P(Z_j=g)}{P(Y_j)}$ où $P(Y_{ij}|Z_j = g)P(Z_j = g)$ se calcule en sommant sur les valeurs de X_{ij} le terme $P(X_{ij} = k, Y_{ij}|Z_j = g)$ obtenu précédemment.
L'algorithme est implémenté dans le logiciel Latent Gold (Vermunt & Magidson, 2015) et permet d'obtenir les estimations des différents paramètres : la taille des clusters, la taille des classes, les paramètres cluster-spécifiques des lois de probabilité spécifiées pour chaque variable (ex : les moyennes cluster-spécifiques pour les variables continues et les probabilités associées aux modalités des variables discrètes), ces mêmes paramètres estimés au niveau des classes, les moyennes (ou probabilités associées au modalités de réponse) des covariables dans chaque cluster, les probabilités d'appartenance aux clusters conditionnellement aux vecteurs des observations, et les probabilités individuelles d'appartenance aux clusters.

La LCCA multi-niveaux présentée ici a le double avantage de prendre en compte la non-indépendance des observations de niveau 1, et de regrouper explicitement les observations de niveau 2 dans des classes qui peuvent décrites et interprétées.

4.2 Modèles mixtes à classes latentes

Cette section présente les modèles mixtes à classes latentes qui seront utilisés pour modéliser les trajectoires de comportements de jeu des joueurs nouvellement inscrits sur le site de la FDJ. Il s'agit d'un travail actuellement en cours, et de plus, le principe des modèles à classes latentes

a été présenté précédemment ; c'est pour ces deux raisons que les méthodes évoquées dans la présente section sont présentées de manière plus succincte.

4.2.1 Principe

Pour modéliser l'évolution d'une variable mesurée à plusieurs temps pour un même sujet, on utilise généralement des modèles mixtes qui permettent de prendre en compte les corrélations entre les mesures d'une même individu grâce à l'introduction d'effets aléatoires. Ce type de modèle fait l'hypothèse que la population est homogène, c'est à dire que les déviations (modélisées par les effets aléatoires) des caractéristiques des individus sont centrées sur des effets fixes communs à tous les individus. On peut cependant supposer dans certains cas que la population est hétérogène et que les paramètres du modèle, en particulier les effets fixes, sont différents selon les sous-populations dont sont issus les individus. Si les appartenances des sujets à ces sous-populations sont inconnues, ces dernières peuvent être définies par des classes latentes.

4.2.2 Écriture des modèles

Modèle linéaire mixte (*growth model*)

Le modèle linéaire mixte "classique" pour données longitudinales (*growth model* en Anglais) s'écrit :

$$Y_{it} = \beta_{0i} + \beta_{1i}t + \epsilon_{it} \quad \text{avec} \quad \begin{cases} \beta_{0i} = \alpha_0 + r_{0i} \\ \beta_{1i} = \alpha_1 + r_{1i} \end{cases} \tag{4.12}$$

où i désigne un individu, t désigne le temps, α_0 est l'intercept, α_1 est la pente, r_{0i} est l'effet aléatoire sur l'intercept, r_{1i} est l'effet aléatoire sur la pente, et ϵ_{it} est le terme d'erreur. De plus, on suppose que les effets aléatoires et le terme d'erreur sont normalement distribués :

$$\begin{cases} r_{0i} \sim \mathcal{N}(0, \sigma_0^2) \\ r_{1i} \sim \mathcal{N}(0, \sigma_1^2) \\ \epsilon_{it} \sim \mathcal{N}(0, \sigma_e^2) \end{cases} \tag{4.13}$$

Modèle de trajectoires basé sur les groupes (*group-based trajectory model*)

Pour prendre en compte l'hétérogénéité de la population et donc l'existence de groupes latents d'individus, (Nagin & Land, 1993) ont proposé un modèle où les différences entre les groupes d'individus sont prises en compte en estimant une trajectoire unique par groupe. Le modèle, appelé *group-based trajectory model*, ou *latent class growth analysis* ou *non-parametric growth model* s'écrit :

$$Y_{it|c_i=k} = \beta_{0k} + \beta_{1k}t + \epsilon_{it} \quad \text{avec} \quad \begin{cases} \beta_{0k} = \alpha_{0k} \\ \beta_{1k} = \alpha_{1k} \end{cases} \tag{4.14}$$

où k désigne une valeur de la variable latente discrète $C = 1, ..., K$ (autrement dit, k désigne une classe latente).
On voit ici que les déviations autour des effets fixes ne sont pas modélisées individuellement mais au niveau des groupes d'individus. C'est à dire que tous les individus d'un groupe k ont exactement la même trajectoire, sans aucune déviation par rapport à β_{0k} et β_{1k}. Ce modèle est parfois appelé *non-parametric growth model* parce que les effets aléatoires sont considérés comme discrets et estimés sans *a priori* sur leur distribution. Malgré la popularité de ce modèle, l'absence de modélisation de la variation inter-individus à l'intérieur des classes peut constituer une limite importante.

Modèle mixte à classes latentes (*growth mixture model*)

Introduit par Verbeke et Lesaffre (1996) et Muthén et Shedden (1999), le modèle mixte à classes latentes (*growth mixture model*) permet, contrairement au modèle de Nagin et Land, de prendre en compte les variations individuelles à l'intérieur des classes. Le modèle s'écrit :

$$Y_{it|c_i=k} = \beta_{0ki} + \beta_{1ki}t + \epsilon_{it} \quad \text{avec} \quad \begin{cases} \beta_{0ki} = \alpha_{0k} + r_{0ki} \\ \beta_{1ki} = \alpha_{1k} + r_{1ki} \end{cases} \tag{4.15}$$

Les paramètres α_{0k} et α_{1k} peuvent être vus comme les effets fixes propres aux classes et capturent les différences de trajectoires globales entre les groupes d'individus. Les effets aléatoires r_{0ki} et r_{1ki} sont supposés normalement distribués et correspondent aux déviations autour de ces paramètres et capturent les différences de trajectoires individuelles à l'intérieur des classes.
Le modèle (4.3) peut être paramétré de plusieurs manières et implique de faire des choix sur :

- le nombre de classes K
- l'indépendance des erreurs ϵ_{it} entre les différentes classes
- l'indépendance des erreurs ϵ_{it} au cours du temps
- la structure de la matrice de variance-covariance des effets aléatoires (corrélation ou non entre les effets aléatoires et matrice indépendante ou non entre les classes)
- l'introduction de termes d'évolution non linéaires (effets quadratiques, effets cubiques, *splines*, etc.)

Estimation du modèle mixte à classes latentes

L'estimation du modèle se fait généralement par maximisation de la vraisemblance à l'aide de l'algorithme EM. Comme pour les modèles présentés dans la section 4.1, il est conseillé de faire varier les valeurs initiales des paramètres pour éviter de trouver une solution correspondant à un maximum local. La sélection de modèle se fait de manière similaire (BIC, entropie, interprétation des classes, etc.). Il est aussi possible d'effectuer des tests de Wald pour décider d'inclure

ou non certains effets fixes ou aléatoires. L'adéquation du modèle peut être testée en représentant graphiquement les trajectoires estimées dans chaque classe et les trajectoires observées des individus.
Les logiciels Mplus (Muthen & Muthen, 2017), Latent Gold (Vermunt & Magidson, 2015), et le package R `lcmm` (Proust-Lima, Philipps, & Liquet, 2017) permettent d'estimer ce type de modèle.

Chapitre 5

Caractérisation des comportements de jeu en ligne et repérage des joueurs à risque

Sommaire

5.1 Typologie des comportements de jeu en ligne 66
5.1.1 Choix d'analyse 66
Définitions des indicateurs utilisés 66
Distributions cluster-spécifiques 67
Estimation 68
Choix du nombre de clusters et du nombre de classes 68
5.1.2 Manuscrit : Typology of online lotteries and scratch games gamblers' behaviours : a multilevel latent class cluster analysis applied to player account-based gambling data (accepté pour publication dans *International Journal of Methods in Psychiatric Research*) 68
5.1.3 Principaux résultats 70
5.1.4 Discussion 71
5.2 Analyse longitudinale de l'activité de jeu des nouveaux joueurs . . 73
5.2.1 Méthodologie 74
5.2.2 Stratégie de sélection de modèle 75
5.2.3 Résultats préliminaires 75
Description des classes 79

5.1 Typologie des comportements de jeu en ligne

Cette section présente les résultats de l'application de la LCCA multi-niveaux décrite dans le chapitre précédent aux données d'activité de jeu des joueurs de la FDJ. L'objectif est d'établir une typologie des comportements de jeu sur Internet, à partir des données d'activité de jeu d'individus jouant à des jeux de hasard pur (loteries à tirage différé, loteries à tirage immédiat et jeux de grattage). Plus précisément, on effectue simultanément une classification des comportements de jeu mensuels (niveau 1) et des comportements de jeu globaux des joueurs (niveau 2). Nous présentons dans un premier temps les choix qui ont été faits quant au modèle de classification, puis nous présentons l'article issu de ce travail et discutons ensuite ses forces et ses limites.

5.1.1 Choix d'analyse

Définitions des indicateurs utilisés

Les variables utilisées dans l'analyse sont légèrement différentes de celles décrites dans le *study protocol*. Au final, nous avons retenu dix indicateurs de comportement de jeu :

- montant total d'argent déposé par le joueur sur son compte au cours du mois
- plus gros montant d'argent déposé au cours du mois
- montant total d'argent misé au cours du mois
- nombre de prises de jeu effectuées au cours du mois
- nombre de jours actifs dans le mois (un jour actif est un jour où le joueur a effectué au moins une prise de jeu)
- nombre de jeux différents auxquels le joueur a joué dans le mois
- utilisation de bonus de fidélité au cours du mois (oui/non)
- chasing ("oui" si le joueur a effectué au moins trois versements dans une période glissante de 12 heures ou effectué un versement dans l'heure qui suit une prise de jeu, "non" sinon)
- montant inhabituel d'argent misé au cours du mois
- nombre inhabituel de jours actifs

Les deux derniers indicateurs servent à identifier une activité de jeu différente de celle habituellement observée pour un joueur donné. Par exemple, pour le premier indicateur, l'idée est de comparer le montant total d'argent qu'un joueur a misé au cours du mois m, par rapport à la moyenne du montant total d'argent misé par mois, calculée sur les mois $m-1$, $m-2$ et $m-3$, que l'on appellera la valeur baseline. Si la valeur au mois m est (très) supérieure à la valeur baseline, on considère que les montants d'argent misés sont inhabituels pour le joueur en question, et que cela peut être le signe d'une possible perte de contrôle de la pratique de jeu. Le second indicateur est défini de la même manière, à partir du nombre de jours actifs. La période de "référence" de trois mois a été définie de manière à avoir une période suffisamment longue

pour estimer correctement l'activité habituelle du joueur, et suffisamment courte pour ne pas avoir besoin de trop de recul
Afin de quantifier les différences observées entre la valeur mesurée à m et la valeur baseline, deux solutions ont été initialement envisagées : calculer la différence entre les deux valeurs ou calculer le ratio des deux valeurs. Cependant, les deux méthodes étaient trop sensibles à l'amplitude des valeurs à comparer. En effet, pour un joueur misant 900 euros au mois m et qui misait en moyenne 800 euros, la différence de 100 euros n'a pas le même sens que pour un joueur qui mise 150 et qui avait l'habitude de miser 50 euros. En calculant plutôt le rapport entre les deux valeurs ($900/800 \simeq 1.1$ pour le premier joueur et $150/50 = 3$ pour le second joueur), l'indicateur semble plus pertinent. En revanche, pour des petites valeurs (ex : un joueur qui mise 30 euros alors qu'il a l'habitude de miser 2 euros, donc un ratio de 15), l'indicateur consistant à calculer le ratio ne sera pas pertinent. Nous avons donc utilisé une autre méthode pour quantifier les activités de jeu inhabituelles : la valeur au mois m a été régressée linéairement sur la valeur baseline, en ajoutant un effet aléatoire sur la valeur baseline. On a ensuite calculé, pour chaque observation, le ratio entre le résidu (valeur prédite-valeur observée) et la racine carrée de la variance totale de l'estimation de la valeur prédite. Ainsi, des ratios élevés indiquent que la valeur observée au mois m est "mal prédite" par la valeur baseline et on considère donc que l'activité au mois m est inhabituelle. Au contraire, des ratios faibles indiquent que la valeur observée au mois m est cohérente avec l'activité moyenne du joueur. Les deux variables créées ont ensuite été standardisées (moyenne=0 et écart-type=1) pour être plus facilement interprétables.
Les covariables utilisées pour décrire les clusters étaient l'âge du joueur, le sexe, l'ancienneté du compte, et le pourcentage de prises de jeu qui concernaient des loteries à tirage différé. Cette dernière covariable permet de faire la distinction entre, d'un côté, les loteries à tirage différé et de l'autre, les loteries à tirage immédiat et les jeux à gratter (les loteries à tirage immédiat sont considérés comme plus proches des jeux de grattage que des loteries à tirage différé, étant données les caractéristiques structurelles des types de jeu). Les moyennes des covariables dans chaque cluster ont été estimées en pondérant par les probabilités individuelles d'appartenance des observations aux clusters. Pour les variables ayant été log-transformées avant l'analyse, les variables originales (non transformées) ont également été incluses comme covariables de manière à faciliter l'interprétation des clusters.

Distributions cluster-spécifiques

Nous avons fait l'hypothèse d'indépendance locale et avons donc spécifié une distribution univariée pour chaque indicateur.

- montant total d'argent déposé : loi normale (après transformation logarithmique)
- plus gros montant d'argent déposé : loi normale (après transformation logarithmique)
- montant total d'argent misé : loi normale (après transformation logarithmique)

- nombre de prises de jeu : loi normale
- nombre de jours actifs : loi binomiale négative
- nombre de jeux différents : loi binomiale négative
- utilisation de bonus de fidélité : loi binomiale
- chasing : loi binomiale
- montant inhabituel d'argent misé : loi normale
- nombre inhabituel de jours actifs : loi normale

Pour les deux indicateurs "nombre de jours actifs" et "nombre de jeux différents", une distribution binomiale négative semblait plus appropriée qu'une distribution de Poisson, étant donné la forte dispersion des valeurs observées. Par ailleurs, un excès de zéros était observé pour ces variables. Une distribution de type *zero-inflated negative binomial* aurait donc pu être envisagée, mais le modèle de mélange rend possible l'utilisation d'un cluster particulier pour "capturer" l'excès de zéros (autrement dit, un cluster qui serait caractérisé par des joueurs-mois sans prise de jeu).

Estimation

Pour éviter le problème des maximas locaux, 100 ensembles des paramètres initiaux ont été générés aléatoirement. Pour chaque ensemble de paramètres, 50 itérations de l'algorithme EM étaient effectuées. Puis, pour les 10 solutions avec les meilleures log-vraisemblances, 100 itérations supplémentaires étaient effectuées. Enfin, pour la meilleure solution à ce stade, l'algorithme EM se poursuivait jusqu'à un maximum de 300 itérations avant de basculer vers l'algorithme Newton-Raphson (50 itérations maximum, tolérance sur la somme des valeurs absolues des changements relatifs des paramètres : 1×10^{-8}) lorsque le seuil de tolérance de l'algorithme EM (0,01) était atteint.

Choix du nombre de clusters et du nombre de classes

Afin de choisir le nombre de clusters optimal, nous avons testé des modèles composés de 1 à 10 clusters. La sélection du nombre de clusters était basée sur un compromis entre le BIC, le taux d'erreurs de classification, la stabilité des solutions, et l'interprétabilité des clusters. Une fois le nombre de clusters sélectionné, nous avons testé des modèles avec ce nombre de clusters en faisant varier le nombre de classes de 1 à 7. Le nombre de classes a été choisi en fonction des mêmes critères que ceux utilisés pour le choix du nombre de clusters.

5.1.2 Manuscrit : Typology of online lotteries and scratch games gamblers' behaviours : a multilevel latent class cluster analysis applied to player account-based gambling data (accepté pour publication dans *International Journal of Methods in Psychiatric Research*)

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

Running title: Typology of online lotteries gamblers

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

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

[*] Correspondence:
Bastien Perrot
bastien.perrot@univ-nantes.fr
U1246 SPHERE - IRS2, 22 Boulevard Bénoni Goullin – 44200 Nantes, France
Tel. +33 253009129

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.

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.

Abstract

Objectives: Internet gambling is often considered as a risk factor for gambling problems compared to 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 were analysed.

Results: We identified 7 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, online gambling, problem gambling, online lotteries, latent class cluster analysis.

1. Introduction

The Internet medium is often considered as a risk factor for gambling problem compared to 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) (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, the betting limits set by the user, etc. 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 & Shaffer (2010), LaBrie & Shaffer (2011) and LaPlante, Schumann, LaBrie, & Shaffer (2008) used data from online sport gamblers; LaPlante, Kleschinsky, LaBrie, Nelson, & Shaffer (2009) and Luquiens et al. (2016) used data from online poker gamblers; Dragicevic, Tsogas, & Kudic (2011) and LaBrie, Kaplan, LaPlante, Nelson, & 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 licenses to gambling operators for 3 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, & 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), i.e. 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 which 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 (i.e. normal, binomial, count data, etc.).

In this study, we performed a 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 one 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 (GNEDS) on 25 March 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.

[INSERT Table 1]

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" intra-gambler 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 intra-gambler 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 in 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 sub-populations 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 3 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 3-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; 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.

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 1-class models composed of 1 to 10 clusters. Once the number of clusters was selected, the corresponding model was tested with 1 to 6 classes. A general rule of thumb is to select the model minimizing the negative log-likelihood function augmented by a penalty function which 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

3.2 Model selection

Fit indices of 1- 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 7-cluster solution because models with more than 7 clusters yield to additional clusters characterized by either non-meaningful interpretation or which seemed very close to other clusters. For example, the 8-cluster model was similar to the 7-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 correspond 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 neither in month *m*, nor in month *m+1*, nor in month *m+2*. This, however, makes less sense or at least is not very interesting and concerns very few observations.

[INSERT Table 2]

Table 3 shows the fit indices for 1- to 7-class 7-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 7-class solution did not bring meaningful information, we selected

the 6-class solution, which in addition (comparing to the 5-class solution) allowed observing a class of gamblers having a high probability of being in the cluster 7 every month.

[INSERT Table 3]

3.3 Parameter estimates

The higher part of Table 4 shows the distribution of the 7 monthly gambling behaviours clusters across the 6 classes of gamblers. The middle part shows the estimated means of the indicators of the cluster specific distributions. The lower part shows the parameter estimates of inactive covariates.

[INSERT Table 4]

3.4 Description of clusters of monthly gambling behaviours

Cluster 1 (40% of the monthly gambling behaviours) corresponded to gambling inactivity, that meant 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 to 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.

Cluster 6 (12% of the monthly gambling behaviours) was characterized by a higher gambling activity compared to the first five clusters (87 euros wagered, 8 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" intra-gambler 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 to 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%).

3.5 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 Cluster 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 (Cluster 2 and Cluster 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 Cluster 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, i.e. 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 7 clusters of monthly gambling behaviours which differed with respect to gambling indicators and 6 classes of gamblers which 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 to 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 a 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 since Bruneau et al. (2016) found that experiencing at least one 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 to 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 like 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 the 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, the Cluster 7 is associated to a high proportion of instant lotteries and scratch games compared to delayed lotteries. In a recent study focusing on-site lottery and scratch games gamblers, Brochado, Santos, Oliveira, & 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). Since 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 seems 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) suggest 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 which 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 is 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.

Competing interests MG-B and GC-B declare that the University Hospital of Nantes has received gambling industry (FDJ and PMU) funding in the form of a sponsorship which 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.

Footnotes:

[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 3 money deposits within a 12h period or if we observe a deposit made less than one hour 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.

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

Association, A. P. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Pub.

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.

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

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.

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*.

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. *Cyberpsychol Behav*, *6*(6), 557-568.

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.

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. 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

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

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. *Comput Hum Behav*, *25*(3), 711-717.

LaPlante, Debi 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, Debi 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*, n/a-n/a. 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 multi-level approach. *Energy Policy*, *102*, 73‑87. https://doi.org/10.1016/j.enpol.2016.12.015

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, Jeroen 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, Jeroen K., & Magidson, J. (2002). Latent class cluster analysis. *Applied latent class analysis*, 89–106.

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

Table 1: Description of gamblers and monthly gambling activity

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

[†] Median [range]

[‡] Frequency (%)

[§] Mean ± standard deviation

Table 2: Fit indices of one-level latent class cluster analysis

	Log-likelihood	BIC	Classification errors
1-Cluster	-1583047	3166299	0.000
2-Cluster	-686860	1374118	0.001
3-Cluster	-354053	708697	0.001
4-Cluster	-190415	381615	0.007
5-Cluster	-138138	277253	0.022
6-Cluster	-108257	217684	0.032
7-Cluster	-79677	160718	0.034
8-Cluster	-59335	120227	0.017
9-Cluster	-35028	71805	0.032
10-Cluster	-18347	38638	0.031

Table 3: Fit indices of two-level latent class cluster analysis

	Log-likelihood	BIC	Classification errors
7-Cluster 1-Class	-79677	160718	0,034
7-Cluster 2-Class	-61505	124453	0,030
7-Cluster 3-Class	-53310	108143	0,031
7-Cluster 4-Class	-49312	100226	0,026
7-Cluster 5-Class	-46751	95184	0,018
7-Cluster 6-Class	-46066	93892	0,025
7-Cluster 7-Class	-44947	91735	0,025

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

	Class size	Cluster1	Cluster2	Cluster3	Cluster4	Cluster5	Cluster6	Cluster7
Class1	0.39	0.77	0.11	0.07	0.01	0.03	0.01	0.00
Class2	0.27	0.20	0.28	0.38	0.02	0.11	0.02	0.00
Class3	0.12	0.02	0.10	0.04	0.08	0.69	0.07	0.00
Class4	0.09	0.37	0.05	0.06	0.13	0.10	0.24	0.06
Class5	0.09	0.00	0.00	0.01	0.09	0.04	0.81	0.04
Class6	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

5.1.3 Principaux résultats

Le modèle de classification retenu est composé de 7 clusters de comportements de jeu mensuels (joueurs-mois) et de 6 classes de joueurs.
Le cluster 1 (40% des joueurs-mois) correspond à une absence d'activité de jeu puisque dans ce cluster, aucune mise et aucun versement d'argent ne sont effectués durant le mois.
Le cluster 2 (13%) est caractérisé par une faible activité de jeu (en moyenne : 11 euros misés, 1,8 jours actifs et pas de dépôt d'argent).
Le cluster 3 (14%) est similaire au cluster 2, mais avec un montant total d'argent misé plus élevé (22 euros en moyenne) et des versements d'argents également plus élevés (25 euros).
Le cluster 4 (4%) est caractérisé par une activité de jeu plus importante que pour les clusters 2 et 3 (en moyenne : 34 euros misés, 5,56 jours actifs et 11 prises de jeu), mais aucun dépôt d'argent.
Le cluster 5 (14%) est similaire au cluster 4 en termes d'argent misé (30 euros) et de nombre de jours actifs (5,1). Les dépôts d'argent sont proches de ceux observés dans le cluster 3 (25 euros) et la probabilité de chasing est de 9%. L'indicateur de nombre inhabituel de jours actifs vaut en moyenne 0,54, proche de la valeur observée pour le cluster 6.
Le cluster 6 (12%) est caractérisé par une activité de jeu plus importante que les cinq premiers clusters (en moyenne : 87 euros misés, 8 jours actifs, 65 euros déposés et 19 prises de jeu). Les deux indicateurs d'activité de jeu inhabituelle valent en moyenne 0,37 pour l'argent misé et 0,58 pour le nombre de jours actifs. La probabilité de chasing est de 20%.
Le cluster 7 (3%) est caractérisé par une très forte activité de jeu comparé aux autres clusters (en moyenne : 483 euros misés, 12,5 jours actifs, 241 euros déposés et 208 prises de jeu). Les deux indicateurs d'activité de jeu inhabituelle valent en moyenne 2,07 pour l'argent misé et 1,11 pour le nombre de jours actifs. La probabilité de chasing est de 61%. La diversité des jeux est plus importante (12 jeux différents en moyenne) que dans les autres clusters, tout comme l'utilisation des bonus (52%). Le pourcentage de loteries à tirage différé est de 19%, ce qui signifie que 81% des prises de jeu concernaient des jeux de grattage ou des loteries à tirage immédiat. Il s'agit enfin du cluster avec le plus haut pourcentage de femmes (46%).

La classe 1 (39% des joueurs) est composée de joueurs qui ont des probabilités de 77%, 11% et 7% d'être, chaque mois, dans les clusters 1, 2 et 3, respectivement. Cela signifie que 39% des individus de l'échantillon ne jouent que très rarement.
Les joueurs de la classe 2 (27%) ont des probabilités de 38%, 28% et 20% d'être dans les clusters 3, 2 et 1, respectivement. Ce sont donc probablement des joueurs occasionnels.
Les joueurs de la classe 3 (12%) ont 69% de chance d'être dans le cluster 5 et de faibles probabilités (entre 0 et 10%) d'être dans les autres clusters. Ils correspondent donc vraisemblablement à des joueurs avec une activité de jeu modérée et stable au cours du temps.
La classe 4 (9%) est composée de joueurs qui ont des probabilités de 37%, 24% et 13% d'être

dans les clusters 1, 6 et 4, respectivement. Ce sont donc des individus qui semblent transiter entre des périodes d'inactivité de jeu (cluster 1) et des périodes caractérisées par une activité de jeu plus importante (clusters 4 et 6).
Les joueurs de la classe 5 (9%) ont 81% de chance d'être dans le cluster 6. Ils semblent donc avoir une forte activité de jeu, stable au cours du temps.
La classe 6 (3%) est composée de joueurs qui ont une probabilité de 81% d'être dans le cluster 7 et 13% de chance d'être dans le cluster 6. Ils semblent donc correspondre à des joueurs avec un comportement de jeu risqué, caractérisé par beaucoup de prises de jeu, une grande variété de jeux, une forte probabilité de chasing et une activité de jeu parfois incohérente avec leur pratique moyenne habituelle. La proportion de femmes (48%) est plus élevée que dans les autres classes de joueurs.

5.1.4 Discussion

L'interprétation des résultats est discutée dans l'article. Si on s'intéresse en particulier à l'identification de comportements de jeu à risque, on peut penser que le cluster 7 (3% des comportements de jeu mensuels) correspond à ce type de comportements de jeu pour plusieurs raisons : une très forte activité de jeu (beaucoup d'argent misé, beaucoup de prises de jeu, beaucoup de jours actifs), une importante probabilité de chasing, des déviations importantes par rapport aux comportements de jeu habituels, et un haut niveau d'implication dans le jeu, telle que mesurée par le nombre de jeux différents pratiqués. Cette forme d'implication dans le jeu semble en effet associée aux problèmes de jeu (LaPlante et al., 2014). De plus, le cluster 7 est caractérisé par une forte proportion de jeux à résultats immédiat, qui permettent des événements de jeu très fréquents. Ce type de jeu semble associé à une pratique problématique des JHA, par rapport aux jeux avec des événements moins fréquents comme les loteries à tirage différé (Challet-Bouju et al., 2015 ; Husky, Michel, Richard, Guignard, & Beck, 2015 ; McCormack et al., 2013).
Les joueurs de la classe 6 (3% des joueurs) ont 81% de chance d'être, chaque mois, dans ce cluster (ou, dit d'une autre manière, ils sont dans le cluster 7 pour 81% des mois). De plus, ils ont de très faibles probabilités d'être dans les clusters caractérisés par une absence (cluster 1) ou une faible activité de jeu (clusters 2 et 3). Ce sont donc des joueurs qui n'ont probablement pas de période d'abstinence de jeu, or l'arrêt de la pratique des JHA pendant au moins un mois apparait comme un facteur protecteur des problèmes liés au jeu (Bruneau et al., 2016). On peut donc penser que les joueurs de la classe 6 sont des "joueurs à risque". Enfin, alors que l'échantillon est composé de 32% de femmes, la proportion de femmes dans la classe 6 est de 48%. Ce résultat peut sembler *a priori* surprenant puisque, d'après les données de la littérature, les hommes semblent plus à risque de problèmes de jeu que les femmes (cf. section 1.2.2). Cependant, les femmes semblent avoir une préférence plus prononcée pour les jeux de grattage que pour les loteries différées, contrairement aux hommes (Brochado, Santos, Oliveira,

& Esperança, 2018). On peut de plus supposer que les femmes ayant des problèmes de jeu ont une préférence pour le jeu sur Internet, qui permet de jouer en évitant le regard des autres et le sentiment de stigmatisation (Griffiths, 2003). Il a aussi été montré que comparativement aux hommes, les femmes avec des problèmes de jeu étaient plus anxieuses et avaient tendance à éviter les activités sociales (Corney & Davis, 2010), ce qui rendrait le jeu sur Internet plus attractif pour cette population.

Concernant le choix du nombre de clusters, les deux critères statistiques – le BIC et le taux d'erreurs de classification – n'ont pas permis de déterminer clairement le nombre de clusters optimal. En effet, les taux d'erreurs de classification étaient très bas pour les modèles de 1 à 10 clusters ($< 3\%$) et le BIC diminuait systématiquement lorsque le nombre de clusters augmentait. Ce dernier problème a été relevé par plusieurs auteurs et semble se produire dans le cadre d'échantillons de grande taille (Paas, 2014 ; Thøgersen, 2017). Le phénomène pourrait aussi être dû à une mauvaise spécification des distributions des variables (Bauer & Curran, 2003). Une solution consiste à ne plus chercher le modèle avec le BIC mais à chercher le modèle à partir duquel le BIC "diminue moins rapidement" ou à regarder les évolutions relatives du BIC. Par exemple, dans son étude, Thøgersen (2017) a considéré qu'une diminution du BIC inférieure à 1% entre un modèle à k clusters et un modèle à $k + 1$ clusters était négligeable et que donc le modèle composé de k clusters était meilleur. Une autre solution est de spécifier des dépendances locales entre certaines paires de variables (Hagenaars, 1988). En effet, le BIC pourrait avoir tendance à privilégier des modèles avec beaucoup de clusters jusqu'à satisfaire "totalement" l'hypothèse d'indépendance locale. Les clusters "supplémentaires" utilisés pour prendre en compte les dépendances locales n'auraient en plus pas forcément été substantiels (Oberski, 2016). En spécifiant explicitement des dépendances locales, on pourrait donc obtenir un modèle avec un nombre plus limité de clusters susbantifs et un modèle plus parcimonieux (puisque la modélisation d'une dépendance locale entre deux variables dans un cluster revient à ajouter un seul paramètre, alors que l'ajout d'un cluster implique l'estimation d'un plus grand nombre de paramètres). La spécification de dépendances locales peut être guidée par l'examination des résidus bivariés (Magidson & Vermunt, 2004) qui permettent de repérer les paires de variables dont la relation est plus forte que celle censée être expliquée par le modèle.
Dans notre cas, les résidus bivariés indiquaient des dépendances locales importantes entre plusieurs paires de variables, notamment entre le montant total d'argent versé sur le compte et le total d'argent misé ou entre le nombre de prises de jeu et l'argent misé. Ces corrélations résiduelles ne sont pas très surprenantes puisqu'on peut logiquement s'attendre à ce que le montant total d'argent déposé sur le compte soit corrélé avec le montant total d'argent misé, et il est peu probable qu'il existe par exemple un cluster dans lequel les joueurs déposent de fortes sommes d'argent sans le dépenser. Nous avons donc spécifié des dépendances locales une par une, en partant des plus importantes selon les résidus bivariés. Après l'ajout de trois

dépendances locales, le BIC continuait tout de même de diminuer jusqu'au modèle le plus complexe testé (10 clusters). La spécification d'une quatrième dépendance locale entrainait des problèmes de convergence du modèle. Au final, nous avons donc choisi de ne pas spécifier de dépendances locales et avons sélectionné le modèle à 7 clusters qui paraissait le plus pertinent sur le plan de l'interprétation, les clusters additionnels des modèles plus complexes semblant ne pas être substantiels.

5.2 Analyse longitudinale de l'activité de jeu des nouveaux joueurs

Le repérage des joueurs à risque de développer des problèmes de jeu est d'autant plus intéressant en matière de prévention qu'il est effectué tôt dans le parcours des joueurs. En effet, les joueurs ayant déjà perdu le contrôle de leur pratique de jeu pourraient être moins réceptifs à des mesures de protection ou chercher à éviter ces dernières en changeant d'opérateur de jeu (Haefeli, Lischer, & Schwarz, 2011 ; Blaszczynski & Nower, 2008). De plus, la sévérité des problèmes de jeu apparait comme un prédicteur du succès du traitement des troubles liés aux JHA (Merkouris, Thomas, Browning, & Dowling, 2016). L'analyse des comportements de jeu de joueurs nouvellement inscrits sur un site de jeu en ligne pourrait permettre de repérer de manière précoce des groupes de joueurs à risque ou commençant à avoir une pratique de jeu problématique.
Dans cette optique, Braverman et Shaffer (2010) ont effectué une classification de 530 joueurs pratiquant des paris sportifs en direct (*live bets*), à partir de l'analyse des données d'activité de jeu pendant le premier mois suivant l'inscription sur l'opérateur de jeu Bwin. Les indicateurs utilisés étaient la fréquence de jeu (mesurée par le nombre de jours actifs), l'intensité de jeu (mesurée par le nombre moyen de paris par jour), la trajectoire de jeu (mesurée par la tendance à diminuer ou augmenter les montants d'argent misés), et la variabilité de jeu (mesurée par l'écart-type des mises d'argent). Les auteurs ont identifié quatre clusters (avec la méthode des *k-means*) et ont calculé dans chacun d'entre eux la proportion de joueurs ayant fermé leur compte en raison de problèmes de jeu. Dans le cluster caractérisé par une forte intensité de jeu et une haute variabilité des mises (3% des joueurs), la proportion de joueurs ayant fermé leur compte en raison de problèmes de jeu était de 73%, contre 45% dans le cluster "faible activité de jeu", 29% dans le cluster "forte activité, faible variabilité", et 32% dans le cluster "activité modérée".
Une des rares études basées sur l'analyse longitudinale des comportements réels de jeu est celle réalisée par Xuan et Shaffer (2009) qui ont analysé plusieurs indicateurs d'activité de jeu (indépendamment les uns des autres) au cours du temps, pour étudier les différences d'évolution entre les joueurs ayant fermé leur compte pour des raison liées à des problèmes de jeu et des joueurs témoins n'ayant pas fermé leur compte. Un des résultats intéressants de cette étude était la tendance des joueurs considérés comme problématiques à effectuer des paris moins risqués (i.e. avec des cotes plus faibles) pour, selon les auteurs, tenter récupérer l'argent perdu.

Nous proposons ici de modéliser l'activité de jeu des joueurs dans les mois suivant leur inscription sur le site de la FDJ. L'objectif est de voir s'il est possible de définir des classes de joueurs avec des trajectoires distinctes, et de caractériser ces classes le cas échéant. Les probabilités d'appartenance à ces classes pourraient ensuite servir à prédire des problèmes de jeu futurs.

5.2.1 Méthodologie

S'agissant d'un travail en préparation, la description des méthodes ci-dessous inclut des choix qui ont été faits récemment et d'autres qui sont encore sujets à des réflexions.
Nous avons choisi d'analyser les données d'activité de jeu pendant les 6 mois suivant la création du compte. Cette période devrait permettre d'observer des comportements de jeu représentatifs, "gommant" l'effet d'adaptation des joueurs qui peut prendre plusieurs semaines ou plusieurs mois. Quatre indicateurs ont été sélectionnés : d'une part, le montant total des mises et le nombre de jours actifs pour représenter l'"intensité" de l'activité de jeu ; d'autre part, le chasing et le nombre de jeux différents pour représenter la "prise de risque" (le nombre de jeux de différents représentant lui même l'implication dans le jeu). L'évolution dans le temps de chaque indicateur sera modélisée à l'aide d'un modèle linéaire mixte à classes latentes, tel que défini dans la section 4.2.2. Il est ensuite prévu de réaliser une analyse en classes latentes à partir des probabilités d'appartenances aux classes obtenues avec le modèle mixte, de manière à définir des groupes de joueurs avec des profils d'intensité de jeu et de prise de risques différents. Il s'agit d'une méthodologie similaire à celle utilisée par Brédart et al. (2016). Une deuxième solution qui pourrait être envisagée est d'utiliser un modèle mixte multivarié à processus latent (Proust-Lima, Amieva, & Jacqmin-Gadda, 2013) qui fait l'hypothèse que toutes les variables mesurent un même phénomène (non observable directement, d'où le terme "latent"). La fonction `multlcmm` (Proust-Lima et al., 2017) permet d'estimer ce type de modèle en l'étendant en plus à une population hétérogène, caractérisée par des classes latentes. Cependant, dans notre cas, l'hypothèse d'existence d'un processus latent commun aux quatre indicateurs est discutable. En effet, comme expliqué plus haut, l'argent dépensé et le nombre de jours actifs correspondent à un concept ("intensité de jeu") différent de celui représenté par le chasing et le nombre de jeux ("prise de risque"). Enfin, une troisième possibilité, qui ne repose pas sur les modèles mixtes, serait d'utiliser le package R `KmL3D` (Genolini et al., 2013) qui adapte l'algorithme des *k-means* pour effectuer une classification de trajectoires basées sur l'analyse conjointe de plusieurs variables longitudinales. Bien que la première méthode envisagée ait le défaut de ne pas modéliser explicitement les relations entres les différents indicateurs, elle apparait plus simple à mettre en œuvre et constitue donc la méthode qui sera utilisée.
Les temps de mesure ont été définis par des périodes de quinze jours, à partir du jour de création du compte. C'est à dire que pour un joueur ayant créé son compte le 5 janvier, le premier temps de mesure t_0 est le 20 janvier, et la valeur de l'indicateur "argent total misé" à t_0 est donc la somme des mises effectuées entre le 5 janvier et le 20 janvier. De la même manière, la valeur

de l'indicateur à t_1 pour ce joueur est la somme des mises effectuées entre le 21 janvier et le 4 février, et ainsi de suite, jusqu'à t_{11}. Les temps de mesure sont donc équidistants et toutes les données des joueurs sont complètes puisqu'on inclut uniquement les joueurs ayant créé leur compte entre le 1er septembre 2015 et le 29 février 2016, c'est à dire ceux dont on dispose des données d'activité de jeu à partir du jour de l'inscription et pendant les 180 jours suivants.

5.2.2 Stratégie de sélection de modèle

Pour chacun des quatre indicateurs, nous avons adopté une stratégie consistant à partir d'un modèle complexe et à le simplifier si nécessaire en fonction des différents critères (convergence et reproductibilité du modèle, BIC, significativité des paramètres, etc.). Pour un nombre de classes fixé, le modèle le plus complexe était défini par un intercept, des effets temps linéaire, quadratique et cubique, et des effets aléatoires sur l'intercept et la pente. Les effets aléatoires étaient supposés non corrélés mais différents entre les classes. Les erreurs résiduelles étaient supposées indépendantes entre les temps de mesure mais différentes entre les classes.
Lorsque le modèle ne convergeait pas correctement ou qu'il était mal répliqué, la variance de l'effet aléatoire sur la pente était fixée à une valeur commune pour toutes les classes. Il en était fait de même pour l'effet aléatoire sur l'intercept si le problème persistait. Si nécessaire, les effets aléatoires sur la pente étaient fixés à zéro dans une ou plusieurs classes. Toujours en cas de problème d'estimation du modèle, les effets fixes sur la fonction cubique du temps étaient supprimés. Ensuite, les effets fixes ou aléatoires non significatifs pour une classe donnée étaient supprimés. Pour chaque modèle testé, nous avons généré aléatoirement 400 ensembles de paramètres initiaux, pour lesquels 10 itérations étaient effectuées. Pour les 50 meilleures solutions à ce stade, l'algorithme d'estimation était effectué entièrement. On considérait que le modèle était "stable" si la log-vraisemblance était répliquée, c'est à dire si au moins deux solutions avait la même log-vraisemblance finale.
La procédure était répétée en faisant varier le nombre de classes, et le modèle final était sélectionné en comparant les modèles retenus pour chaque valeur du nombre de classes. Les analyses ont été effectuées avec le logiciel Mplus Version 8.1 (Muthen & Muthen, 2017).

5.2.3 Résultats préliminaires

Le tableau 5.1 présente les caractéristiques et les propriétés statistiques des modèles composés de 1 à 8 classes.

TABLE 5.1 – Propriétés statistiques des modèles mixtes à classes latentes composés de 1 à 8 classes

K	LL	BIC	p	entropie	spécification
1	-22428	44905	7	1	[i s q c] / i* / s*
2	-21369	42843	15	0,8	[i s q c] / i* / s*
3	-20955	42043	19	0,9	[i s q c] / i / s
4	-20717	41611	25	0,83	[i s q c] / i / s
5	-20588	41395	31	0,81	[i s q c] / i / s
6	-20469	41193	37	0,86	[i s q c] / i / s
7	-20364	41023	42	0,83	[i s q c] / i
8	non répliquée	41791	40	0,8	[i s q] / i

K : nombre de classes ; LL : log-vraisemblance ; p : nombre de paramètres ; [i s q c] signifie que le modèle contient un intercept [i], une fonction linéaire du temps [s], une fonction quadratique du temps [q], et une fonction cubique du temps [c] ; i signifie que le modèle possède un effet aléatoire sur l'intercept ; s signifie que le modèle possède un effet aléatoire sur la pente ; les notations i* et s* signifient que la variance de l'effet aléatoire sur l'intercept et sur la pente varie, respectivement, en fonction des classes.

Le BIC diminuait avec le nombre de classes sans atteindre un minimum. Le BIC du modèle à 7 classes était plus faible que celui du modèle à 8 classes, mais la vraisemblance de ce dernier n'a pas était répliquée. Nous avons donc choisi le modèle composé de 7 classes. Le tableau 5.2 indique les probabilités moyennes d'appartenance aux classes en fonction de la classe la plus probable (affectation modale) pour le modèle à 7 classes. Les valeurs proches de 100% sur la diagonale indiquent que le modèle discrimine bien les individus qui sont dans des classes différentes. La taille relative des classes est indiquée sur la première ligne du tableau.

TABLE 5.2 – Probabilités moyennes d'appartenance aux classes en fonction de la classe la plus probable

	Probabilités moyennes d'appartenance aux classes (%)						
Classe la plus probable	1 (23)	2 (15)	3 (12)	4 (4)	5 (13)	6 (3)	7 (26)
1	84	2	3	0	5	0	7
2	3	93	0	0	4	0	0
3	4	0	82	0	2	1	12
4	0	0	0	97	0	0	2
5	9	2	3	0	85	0	1
6	0	0	2	1	0	91	6
7	4	0	7	0	0	1	88

Exemple d'interprétation : les joueurs affectées à la classe selon l'affectation modale ont des probabilités de 84% d'appartenir à la classe 1 (dont la taille relative est de 23%), 2% d'appartenir à la classe 2 (taille relative : 15%), 3% d'appartenir à la classe 3 (taille relative : 12%), etc.

Les estimations des effets fixes du modèle à sept classes latentes sont présentées dans le tableau 5.3.

TABLE 5.3 – Estimations des effets fixes du modèle mixte à sept classes latentes

Paramètre	Estimation	Erreur-type	p-value
Classe 1			
Intercept	1.741	0.183	<0.001
Effet linéaire temps	-0.480	0.093	<0.001
Effet quadratique temps	0.065	0.016	<0.001
Effet cubique temps	-0.003	0.001	<0.001
Classe 2			
Intercept	1.979	0.081	<0.001
Effet linéaire temps	-1.146	0.047	<0.001
Effet quadratique temps	0.189	0.008	<0.001
Effet cubique temps	-0.009	0.000	<0.001
Classe 3			
Intercept	3.554	0.189	<0.001
Effet linéaire temps	-0.189	0.184	0.304
Effet quadratique temps	-0.055	0.034	0.100
Effet cubique temps	0.004	0.002	0.014
Classe 4			
Intercept	3.074	0.103	<0.001
Effet linéaire temps	-0.104	0.039	0.008
Effet quadratique temps	0.020	0.008	0.010
Effet cubique temps	-0.001	0.000	0.007
Classe 5			
Intercept	3.487	0.126	<0.001
Effet linéaire temps	-1.547	0.113	<0.001
Effet quadratique temps	0.219	0.023	<0.001
Effet cubique temps	-0.010	0.001	<0.001
Classe 6			
Intercept	4.266	0.839	<0.001
Effet linéaire temps	0.077	0.197	0.695
Effet quadratique temps	0.007	0.047	0.878
Effet cubique temps	-0.002	0.003	0.566
Classe 7			
Intercept	2.524	0.088	<0.001
Effet linéaire temps	-0.491	0.088	<0.001
Effet quadratique temps	0.092	0.015	<0.001
Effet cubique temps	-0.005	0.001	<0.001

La figure 5.1 représente les trajectoires moyennes estimées de l'argent misé (log-euros) au

cours du temps, pour chacune des sept classes définies par le modèle.

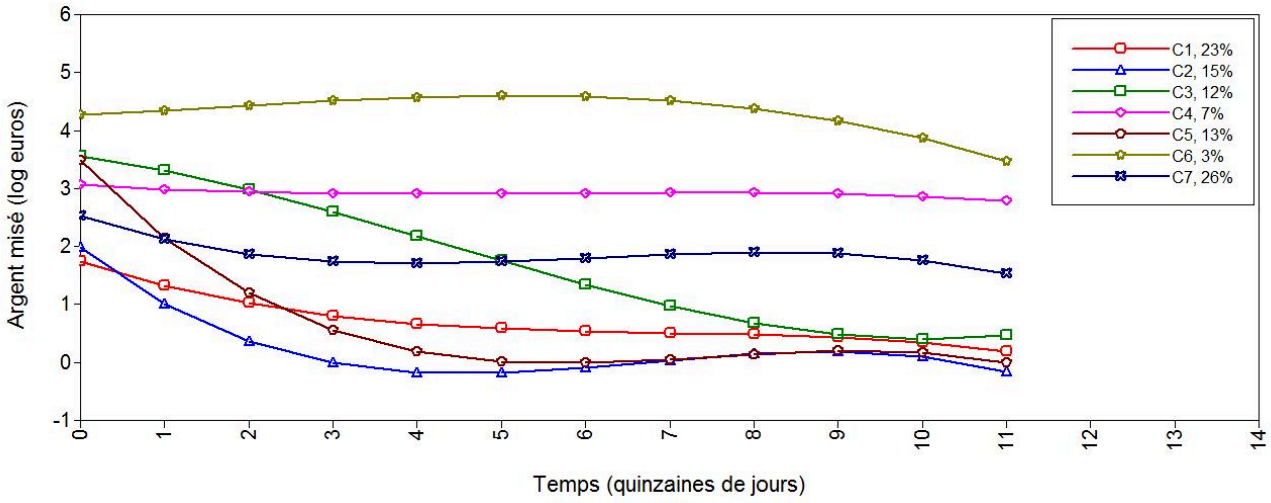


FIGURE 5.1 – Trajectoires moyennes estimées de l'argent misé (log euros) pour les sept classes définies par le modèle mixte à classes latentes

La figure 5.2 permet de comparer les trajectoires moyennes observées dans chaque classe et les trajectoires moyennes estimées.

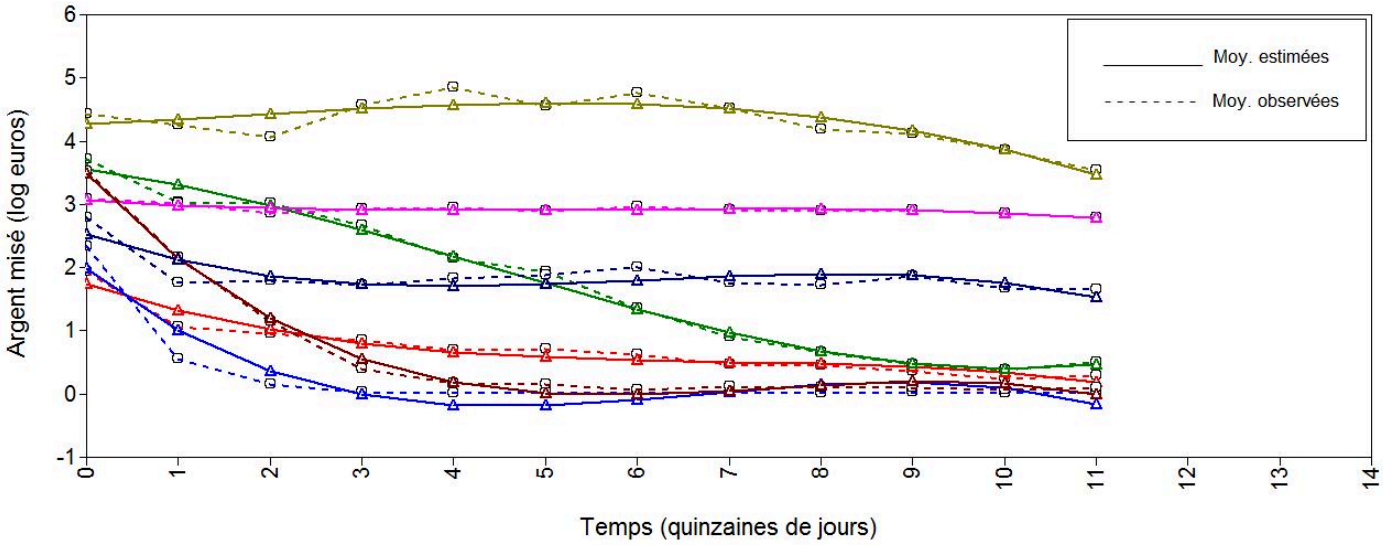


FIGURE 5.2 – Trajectoires moyennes estimées et observées pour les sept classes définies par le modèle

La figure C.-1 en annexe C représente, pour chaque classe, les trajectoires estimées moyennes et les trajectoires individuelles observées.

Description des classes La classe 1 (23% des joueurs) correspond à une activité de jeu faible au départ (1,7 log-euros misés durant les 15 premiers jours, c'est à dire environ 4 euros sur l'échelle originale), et qui décroit légèrement au cours du temps, entre 1 et 3 euros dépensés par période de 15 jours.
Les joueurs de la classe 2 (15% des joueurs) dépensent des sommes similaires à ceux de la classe 1 dans les 15 jours suivant l'inscription, mais ces sommes décroissent plus rapidement jusqu'à 2 mois après l'inscription (T=4), date à partir de laquelle les joueurs de cette classe cessent de miser de l'argent. La figure 5.2 permet de voir que les sommes d'argent misées sont effectivement nulles à partir de cette date, contrairement à ce que peuvent suggérer les moyennes estimées par le modèle qui semblent indiquer une légère augmentation des mises à partir de T=6.
Les joueurs de la classe 5 (13% des joueurs) dépensent au départ plus d'argent que les joueurs des classes 1 et 2, mais, comme pour la classe 2, semblent cesser de jouer 3 mois après leur inscription.
Les joueurs de la classe 3 (12% des joueurs) misent, dans les 15 premiers jours, des sommes d'argent similaires à celles des joueurs de la classe 5, avec une tendance à la baisse relativement linéaire jusqu'à environ T=9, où la moyenne d'argent misé est alors d'environ 1 euros par période de 15 jours.
La classe 7 (26% des joueurs) correspond à une activité de jeu assez faible au départ (environ 11 euros misés durant les 15 premiers jours) qui décroit légèrement et se stabilise autour de 5 euros dépensés par période de 15 jours.
Les joueurs de la classe 4 (7% des joueurs) misent de manière constante environ 19 euros par période de 15 jours, au cours des 6 mois de suivi.
La classe 6 (3% des joueurs) est caractérisée par des sommes d'argent misées sensiblement plus élevées que dans les autres classes. En moyenne, les joueurs de cette classe dépensent environ 89 euros par période de 15 jours, durant les 3 premiers mois, date à partir de laquelle on observe une légère diminution de l'argent misé.

Chapitre 6

Discussion et perspectives

Sommaire

6.1 Discussion 82
6.1.1 Résumé et limites des travaux 82
Développement du RSPG pour le repérage rapide des joueurs problématiques dans les enquêtes à grande échelle 82
Développement du POPS pour évaluer la pratique excessive du poker . 82
Classification des comportements de jeu sur Internet 83
6.1.2 Discussion sur le jeu responsable 83
6.2 Perspectives 84
6.2.1 Poursuite de l'étude EDEIN 84
6.2.2 Méthode de classification 85
6.2.3 Analyse longitudinale des comportements de jeu 85
6.3 Conclusion 86

6.1 Discussion

6.1.1 Résumé et limites des travaux

Les travaux présentés dans cette thèse traitent du repérage des personnes ayant une pratique problématique des JHA. Nous avons dans un premier temps présenté le développement d'un questionnaire permettant un repérage rapide et homogène des pratiques problématiques de jeu dans les études à grande échelle. Dans un deuxième temps, nous avons discuté du cas particulier du poker, dont les caractéristiques structurelles, en particulier la présence d'adresse réelle, semblent en faire un jeu à part parmi l'ensemble des JHA. Dans un troisième temps, nous avons présenté le travail principal de cette thèse, consistant à réaliser une classification des comportements de jeu réels sur Internet, pour tenter d'identifier des comportements à risque. Enfin, nous avons introduit les objectifs et la méthodologie d'une future analyse consistant à identifier des trajectoires de joueurs, en fonction de leur comportement de jeu suivant l'inscription sur le site d'un opérateur de jeu en ligne. Nous rappelons ci-dessous les résultats principaux et les limites de ces travaux.

Développement du RSPG pour le repérage rapide des joueurs problématiques dans les enquêtes à grande échelle

Le RSPG est composé d'un hétéro-questionnaire composé de trois items, destiné à une utilisation par téléphone ou en face à face (sensibilité = 95%, spécificité = 78%) et d'un auto-questionnaire composé de quatre items destiné aux enquêtes en ligne (sensibilité = 95%, spécificité = 64%). Les deux questionnaires permettent une première phase de dépistage relativement performante et économique par rapport aux questionnaires généralement utilisés. Parmi les limites de cet outil, les deux questionnaires finaux n'ont pas été validés sur un échantillon indépendant. De plus, certains items des questionnaires les plus utilisés, notamment l'ICJE n'ont pas pu être inclus dans la liste des items à tester, et la formulation de certains items testés ne correspondaient pas exactement à celle des items originaux. De plus, seule la version française des deux questionnaires a été étudiée.

Développement du POPS pour évaluer la pratique excessive du poker

Le POPS est composé de six items mesurant les conséquences négatives et la perte de contrôle liées à une pratique excessive du poker. Si l'outil développé semble mesurer de manière plus adéquate certaines caractéristiques des troubles liés aux jeu, et semble mieux capturer les problèmes de jeu actuels par rapport à l'ICJE, il présente plusieurs limites. La première est le manque de diagnostic clinique dans l'étude pour valider cliniquement le POPS. La deuxième concerne l'évaluation des propriétés psychométriques du questionnaire ; en particulier, la consistance interne mesurée par l'alpha de Cronbach était inférieure à 0,70 et d'autres propriétés n'ont pas pu être évaluées. Cependant, comme évoqué dans la section 2.1, si le questionnaire est utilisé à

des fins "diagnostiques", les propriétés psychométriques seraient considérées comme secondaires, par rapport aux capacités diagnostiques telles que le sensibilité et la spécificité. En revanche, si le questionnaire a plutôt pour but de "décrire" les problèmes de jeu potentiels (notamment en prenant en compte les deux dimensions définies pour le POPS), l'outil devra présenter des propriétés psychométriques satisfaisantes.

Classification des comportements de jeu sur Internet

La classification basée sur les classes latentes a permis d'identifier sept clusters de comportements de jeu sur des personnes pratiquant les JHA de hasard pur sur Internet. Un de ces clusters est caractérisé notamment par une très forte activité de jeu, un comportement de chasing fréquent et une forte implication dans le jeu, ce qui nous fait penser qu'il s'agit d'un cluster regroupant des comportements de jeu à risque. Un autre résultat intéressant est la forte proportion de femmes (48%) parmi les joueurs ayant ce type de comportement de jeu. Les limites de ce travail concernent en premier lieu la sélection du modèle final, notamment le choix du nombre de clusters. En effet, les critères statistiques n'indiquaient pas précisément qu'un modèle était vraiment meilleur que les autres, et l'interprétation des clusters a donc orienté le choix du modèle final. La solution consistant à ajouter des dépendances locales dans le modèle ne s'est pas avérée concluante. Par ailleurs, le choix des variables sélectionnées pour l'analyse pourrait également être discuté.
Au niveau de l'interprétation des résultats, les données d'activité de jeu ne permettent pas à elles seules d'identifier précisément des comportements à risque. En effet, comme noté par Gainsbury (2011), plusieurs éléments contextuels doivent aussi être pris en compte, tels que le niveau de revenus, l'état d'esprit pendant les sessions de jeu, et bien évidemment les conséquences de la pratique de jeu. On peut également se poser la question du sens de la relation entre les comportements de jeu à risque observés dans l'étude et les troubles liés au jeu : un comportement de jeu inadapté pourrait être une des causes du développement d'un trouble lié à la pratique des JHA, mais le fait qu'un joueur souffre d'une addiction aux JHA pourrait aussi expliquer l'observation d'une pratique de jeu inadaptée (Sharpe, 2002). L'analyse que nous avons effectuée ne permet pas de répondre à cette question.

6.1.2 Discussion sur le jeu responsable

L'application pratique du projet EDEIN est l'implémentation d'un modèle de repérage des problèmes de jeu dans un outil permettant une surveillance en routine de l'activité des joueurs en ligne. Cet outil devra être en mesure d'évaluer, quantitativement et qualitativement, les problèmes de jeu des personnes possédant un compte chez les opérateurs de jeu en ligne autorisés en France. Des mesures de prévention adaptées au profil des joueurs pourront ensuite être appliquées. Ce système s'inscrit donc dans une politique de jeu responsable et pose plusieurs questions d'ordre technique et éthique.

L'un des problèmes évoqués par les chercheurs concerne l'inquiétude que peut susciter l'analyse des données personnelles d'activité de jeu des joueurs. S'il est compréhensible (dans une logique de maximisation des profits) que les opérateurs de jeu cherchent à utiliser ces données pour inciter leurs clients à jouer, Griffiths et Parke (2002) notent que les opérateurs les moins scrupuleux pourraient chercher à identifier volontairement les joueurs problématiques et en faire une cible marketing privilégiée. Sans aller aussi loin, les problèmes liés au paradoxe entre les contraintes économiques d'un opérateur de jeu et l'application d'une politique de jeu responsable (obligatoire en France) peuvent être illustrés par la part du chiffre d'affaires des opérateurs attribuable aux joueurs problématiques. En France, Costes (2015) a estimé qu'en 2014, 40% du chiffre d'affaires des opérateurs de jeu (57% pour les jeux en ligne) était attribuable aux joueurs à risque modéré (ICJE ≥ 3) ou problématiques (ICJE ≥ 8), alors qu'ils représentaient 4,7% des joueurs. De manière similaire, Hancock, Schellinck, et Schrans (2008) estimaient que cette part variait entre 30% et 50%. Ces chiffres donnent une indication des conséquences qu'une protection efficace des joueurs vulnérables pourrait avoir sur l'activité économique d'un opérateur de jeu. Cela pose aussi la question de la responsabilité sociétale (ou sociale) des entreprises proposant des produits ou des activités potentiellement dangereuses, telles que les JHA. A ce propos, Yani-de Soriano, Javed, et Yousafzai (2012) écrivent que, d'un point de vue éthique, le financement des engagements sociaux des opérateurs de jeu, même philanthropiques (ex : dons à des œuvres caritatives, organisation d'événements solidaires, etc.) ne devraient pas reposer sur les revenus provenant de joueurs problématiques (environ la moitié des revenus), et qu'ils devraient au contraire aider ces joueurs par la mise en place de mesures de prévention. On peut ajouter que la détection et les actions d'information, de prévention et de soins qui s'en suivent devraient être gérées par des structures de soins spécialisées, et pas seulement par les opérateurs.

6.2 Perspectives

6.2.1 Poursuite de l'étude EDEIN

L'analyse des comportements de jeu réels que nous avons effectuée portait uniquement sur les données d'un seul opérateur de jeu en ligne (la FDJ) et uniquement sur les jeux de hasard pur. Un accès futur aux données de l'ARJEL, comme prévu dans le cadre de l'étude EDEIN pourrait permettre d'étudier les comportements de jeu pour tous les types de JHA autorisés en France. Cela permettrait en plus d'observer l'activité de jeu complète des joueurs sur tous les sites des opérateurs de jeu agréés sur lesquels les joueurs possèdent un compte.
La deuxième phase du projet EDEIN consiste à croiser les données de comportement de jeu aux réponses à l'ICJE récoltées parmi un nouvel échantillon de joueurs. A partir de l'ensemble de ces données, nous serons en mesure de développer un modèle prédictif du statut des joueurs défini par l'ICJE. Les résultats du modèle pourront être comparés au statut des joueurs défini par la classification basée sur les classes latentes. Les entretiens de la phase 3 serviront à valider clini-

quement le modèle prédictif, étudier la pertinence du modèle de classification, et potentiellement de trouver des nouveaux indicateurs de comportement de jeu, en particulier des indicateurs de chasing ou de prise de risque. Cette phase permettra également d'avoir accès à des éléments contextuels, comme les dommages liés à la pratique de jeu ou le niveau de revenus, et ainsi envisager des actions plus personnalisées.

6.2.2 Méthode de classification

Puisqu'il est prévu de réaliser une typologie des comportements de jeu à partir des données de l'ARJEL, cela pourrait constituer une opportunité pour appliquer une méthode de classification différente de celle utilisée dans cette thèse. En effet, une des limites notée concernait le choix du nombre de clusters, notamment en se basant sur des critères statistiques tels que le BIC. Dans le cas où un grand nombre de clusters semble nécessaire pour obtenir une bonne adéquation du modèle, van den Bergh, Schmittmann, et Vermunt (2017) ont proposé un modèle de classification basée sur les classes latentes avec une division hiérarchique des clusters (*Latent Class Tree*). L'idée globale est de séparer l'échantillon initial X en 2 clusters X_1 et X_2 à l'aide d'un modèle à classes latentes. Si le BIC indique que le modèle à 2 classes (avec X_1 et X_2) est meilleur que le modèle à 1 classe (avec seulement X), on applique un modèle à 2 classes latentes sur X_1 pour obtenir deux sous-échantillons X_{11} et X_{12}. Cette subdivision est conservée si le BIC du modèle à 2 classes (avec X_{11} et X_{12}) est meilleur que le modèle à 1 classe (avec uniquement X_1). On applique la même méthode au sous-échantillon X_2 pour tester la partition composée de X_{21} et X_{22}. On continue ainsi jusqu'à ce que le BIC indique que plus aucune subdivision n'est nécessaire. Dans notre cas, X_1 serait par exemple un cluster caractérisé par une inactivité de jeu (comme notre cluster 1 qui regroupe 40% des observations) et X_2 correspondrait alors à un cluster composé de 60% des observations, caractérisé par une activité de jeu "non-nulle". X_1 ne serait ensuite vraisemblablement pas divisé et on se focaliserait donc sur X_2. Cette méthode permettrait de se focaliser sur les divisions les plus pertinentes et donnerait une idée des relations hiérarchiques entre les clusters, tout en limitant leur nombre.

6.2.3 Analyse longitudinale des comportements de jeu

Une analyse longitudinale des données d'activité de jeu au cours des 12 mois pourrait permettre de répondre en partie à la question précédemment posée sur la relation entre le comportement de jeu et les troubles liés au jeu. Bien qu'il agisse d'un cercle vicieux complexe (Sharpe, 2002), on peut penser que l'étude de l'évolution des comportements de jeu dans le temps pourrait permettre de distinguer les comportements "risqués" qui existent en amont de la perte de contrôle, et les comportements, peut-être plus "extrêmes", qui sont plutôt la conséquence d'une perte de contrôle.

6.3 Conclusion

Le repérage des comportements de jeu excessifs ou à risque constitue une étape importante dans les démarches de prévention des problèmes liés à la pratique des JHA. Si les entretiens cliniques restent la référence pour poser un diagnostic de trouble lié au jeu, deux grandes méthodes peuvent être envisagées pour effectuer un repérage des problèmes de jeu en population générale, ou dans la population des joueurs. La première solution est basée sur l'utilisation de questionnaires, qui permettent d'évaluer la présence et/ou la sévérité de potentiels problèmes de jeu. Nous avons décrit les avantages et les inconvénients de ces questionnaires, en particulier leur adéquation aux JHA faisant intervenir une part d'adresse, et avons notamment proposé un outil d'évaluation de la pratique excessive du poker. La seconde méthode est basée sur l'observation des comportements de jeu réels, définis par l'analyse d'indicateurs d'activité de jeu. Dans le contexte des JHA sur Internet, nous avons montré que les données recueillies par les opérateurs de jeu pouvaient permettre de définir des comportements de jeu distincts, et notamment des comportements de jeu pouvant être considérés comme problématiques. De plus, grâce à l'utilisation d'une méthode statistique peu utilisée dans ce contexte, une analyse en classes latentes multi-niveaux, nous avons également pu identifier des profils de joueurs caractérisés par des comportements de jeu à risque. La combinaison des données subjectives rapportées par les joueurs et des données objectives d'activité de jeu permettrait de développer un modèle de repérage des comportements de jeu à risque et des joueurs à risque ou problématiques, et de mettre en place des actions de prévention adaptées au profil des joueurs.

Références

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. doi: 10.1080/14459795.2012.754919

ARJEL. (2013). *Lutter contre le jeu excessif ou pathologique. Recommandations trois ans après l'adoption de la loi d'ouverture du marché des jeux en ligne* (Rapport).

Auer, M., & Griffiths, M. D. (2017). Self-Reported Losses Versus Actual Losses in Online Gambling : An Empirical Study. *Journal of Gambling Studies*, *33*(3), 795–806. doi: 10.1007/s10899-016-9648-0

Baggio, S., Dupuis, M., Berchtold, A., Spilka, S., Simon, O., & Studer, J. (2017). Is gambling involvement a confounding variable for the relationship between Internet gambling and gambling problem severity ? *Computers in Human Behavior*, *71*, 148–152. doi: 10.1016/j.chb.2017.02.004

Banfield, J. D., & Raftery, A. E. (1993). Model-Based Gaussian and Non-Gaussian Clustering. *Biometrics*, *49*(3), 803–821. doi: 10.2307/2532201

Barrault, S., Untas, A., & Varescon, I. (2014). Special features of poker. *International Gambling Studies*, *14*(3), 492–504. doi: 10.1080/14459795.2014.968184

Barrault, S., & Varescon, I. (2013). Impulsive sensation seeking and gambling practice among a sample of online poker players : Comparison between non pathological, problem and pathological gamblers. *Personality and Individual Differences*, *55*(5), 502–507. doi: 10.1016/j.paid.2013.04.022

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. doi: 10.1037/1082-989X.8.3.338

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the Science of Self-Reports and Finger Movements : Whatever Happened to Actual Behavior ? *Perspectives on Psychological Science*, *2*(4), 396–403. doi: 10.1111/j.1745-6916.2007.00051.x

Bijmolt, T. H., Paas, L. J., & Vermunt, J. K. (2004). Country and consumer segmentation : multi-level latent class analysis of financial product ownership. *International Journal of Research in Marketing*, *21*(4), 323–340. doi: 10.1016/j.ijresmar.2004.06.002

Binde, P. (2011). What are the most harmful forms of gambling ? Analyzing problem gambling prevalence surveys. *rapport nr. : CEFOS Working Papers 12*.

Binde, P., Romild, U., & Volberg, R. A. (2017). Forms of gambling, gambling involvement and problem gambling : evidence from a Swedish population survey. *International Gambling Studies*, *17*(3), 490–507. doi: 10.1080/14459795.2017.1360928

Bjerg, O. (2010). Problem gambling in poker : money, rationality and control in a skill-based social game. *International Gambling Studies*, *10*(3), 239–254. doi: 10.1080/14459795.2010.520330

Blaszczynski, A., Ladouceur, R., & Shaffer, H. J. (2004). A science-based framework for responsible gambling : the Reno model. *Journal of Gambling Studies*, *20*(3), 301–317. doi: 10.1023/B:JOGS.0000040281.49444.e2

Blaszczynski, A., & Nower, L. (2002). A pathways model of problem and pathological gambling. *Addiction (Abingdon, England)*, *97*(5), 487–499.

Blaszczynski, A., & Nower, L. (2008). Differences in attitudes toward money between subgroups of gamblers : Implications for smart card technologies and an exploration of the Tool and Drug Theories of Money in gambling. *Report prepared for the Queensland Government Treasury*.

Bouju, G., Grall-Bronnec, M., Quistrebert-Davanne, V., Hardouin, J.-B., & Vénisse, J.-L. (2013). Texas hold'em poker : a qualitative analysis of gamblers' perceptions. *Journal of Gambling Issues*, *28*, 1–28. doi: 10.4309/jgi.2013.28.4

Boutin, C. (2010). *Le jeu : chance ou stratégie ? : choisir librement la place du jeu dans votre vie*. Éditions de l'Homme.

Braverman, J., LaPlante, D. A., Nelson, S. E., & Shaffer, H. J. (2013). Using cross-game behavioral markers for early identification of high-risk internet gamblers. *Psychology of Addictive Behaviors*, *27*(3), 868–877. doi: 10.1037/a0032818

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. doi: 10.1093/eurpub/ckp232

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. doi: 10.1037/a0036428

Brédart, A., Merdy, O., Sigal-Zafrani, B., Fiszer, C., Dolbeault, S., & Hardouin, J.-B. (2016). Identifying trajectory clusters in breast cancer survivors' supportive care needs, psychosocial difficulties, and resources from the completion of primary treatment to 8 months later. *Supportive Care in Cancer*, *24*(1), 357–366. doi: 10.1007/s00520-015-2799-1

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

Broda, A., LaPlante, D. A., Nelson, S. E., LaBrie, R. A., Bosworth, L. B., & Shaffer, H. J. (2008). Virtual harm reduction efforts for Internet gambling : effects of deposit limits on actual Internet sports gambling behavior. *Harm Reduction Journal*, *5*(1), 27. doi: 10.1186/1477-7517-5-27

Browne, B. R. (1989). Going on tilt : Frequent poker players and control. *Journal of gambling behavior*, *5*(1), 3–21. doi: 10.1007/BF01022134

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. doi: 10.1016/j.addbeh.2016.01.010

Caillon, J., Grall-Bronnec, M., Challet-Bouju, G., & Romo, L. (2017). Revue systématique de la littérature évaluant l'efficacité des outils de jeu responsable existants pour les jeux de hasard et d'argent en ligne : mise au point et perspectives, A systematic review of literature evaluating the effectiveness of existing gaming tools for online gambling : focus and perspectives. *Psychotropes*, *23*(3), 21–40. doi: 10.3917/psyt.233.0021

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.

Caillon, J., Grall-Bronnec, M., Perrot, B., Leboucher, J., Donnio, Y., Romo, L., & Challet-Bouju, G. (2018). Effectiveness of At-Risk Gamblers' Temporary Self-Exclusion from Internet Gambling Sites. *Journal of Gambling Studies*, 1–15. doi: 10.1007/s10899-018-9782-y

Calado, F., & Griffiths, M. D. (2016). Problem gambling worldwide : An update and systematic review of empirical research (2000–2015). *Journal of Behavioral Addictions*, *5*(4), 592–613. doi: 10.1556/2006.5.2016.073

Celeux, G., & Govaert, G. (1995). Gaussian parsimonious clustering models. *Pattern Recognition*, *28*(5), 781–793. doi: 10.1016/0031-3203(94)00125-6

Celeux, G., & Soromenho, G. (1996). An entropy criterion for assessing the number of clusters in a mixture model. *Journal of Classification*, *13*(2), 195–212. doi: 10.1007/BF01246098

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. doi: 10.1007/s10899-014-9496-8

Challet-Bouju, G., Perrot, B., Romo, L., Valleur, M., Magalon, D., Fatséas, M., ... Hardouin, J. (2016). Harmonizing Screening for Gambling Problems in Epidemiological Surveys - Development of the Rapid Screener for Problem Gambling (RSPG). *J Behav Addict*, *5*, 239–50. doi: 10.1556/2006.5.2016.030

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. doi: 10.4309/jgi.2010.24.8

Cortina, J. M. (1993). What is coefficient alpha ? An examination of theory and applications.

Journal of applied psychology, *78*(1), 98.

Costes, J.-M. (2015). *Le « jeu responsable » est-il un concept pertinent pour la prévention des problèmes de jeu ?* Nanterre.

Costes, J.-M., & Eroukmanoff, V. (2018). Les pratiques de jeux d'argent sur Internet en France en 2017. , 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).

Costes, J. M., Pousset, M., Eroukmanoff, V., Le Nezet, O., Richard, J. B., Guignard, R., ... Arwidson, P. (2011). Les niveaux et pratiques des jeux de hasard et d'argent en 2010. *Tendances*, *77*, 1 – 8.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *psychometrika*, *16*(3), 297–334.

Currie, S. R., Hodgins, D. C., & Casey, D. M. (2013). Validity of the Problem Gambling Severity Index Interpretive Categories. *Journal of Gambling Studies*, *29*(2), 311–327. doi: 10.1007/s10899-012-9300-6

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, 1–38.

de Vet, H. C. W., Terwee, C. B., Mokkink, L. B., & Knol, D. L. (2011). *Measurement in Medicine : A Practical Guide*. Cambridge : Cambridge University Press. doi: 10.1017/CBO9780511996214

Devos, G., Challet-Bouju, G., Burnay, J., Maurage, P., Grall-Bronnec, M., & Billieux, J. (2016). Adaptation and validation of the Gambling Motives Questionnaire-Financial (GMQ-F) in a sample of French-speaking gamblers. *International Gambling Studies*, *0*(0), 1–15. doi: 10.1080/14459795.2016.1264080

Dowling, N. A., Merkouris, S. S., Greenwood, C. J., Oldenhof, E., Toumbourou, J. W., & Youssef, G. J. (2017). Early risk and protective factors for problem gambling : A systematic review and meta-analysis of longitudinal studies. *Clinical Psychology Review*, *51*, 109–124. doi: 10.1016/j.cpr.2016.10.008

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. doi: 10.1080/14459795.2011.629204

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. *The European Journal of Health Economics*. doi: 10.1007/s10198-017-0945-z

Eroukmanoff, V., Costes, J. M., & Tovar, M. L. (2014). *Les joueurs de poker, une population présentant un profil particulier ?* Paris : Observatoire des jeux (ODJ).

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster Analysis*. John Wiley & Sons.

Fayers, P. M., & Machin, D. (2013). *Quality of life : the assessment, analysis and interpretation of patient-reported outcomes*. John Wiley & Sons.

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

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of eugenics*, *7*(2), 179–188.

Fraley, C., & Raftery, A. E. (1998). How many clusters ? Which clustering method ? Answers via model-based cluster analysis. *The computer journal*, *41*(8), 578–588.

Fraley, C., & Raftery, A. E. (2002). Model-based clustering, discriminant analysis, and density estimation. *Journal of the American statistical Association*, *97*(458), 611–631.

Fraley, C., & Raftery, A. E. (2003). Enhanced Model-Based Clustering, Density Estimation, and Discriminant Analysis Software : MCLUST. *Journal of Classification*, *20*(2), 263–286. doi: 10.1007/s00357-003-0015-3

Gainsbury, S. M. (2011). Player account-based gambling : potentials for behaviour-based research methodologies. *International Gambling Studies*, *11*(2), 153–171. doi: 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. doi: 10.1007/s40429-015-0057-8

Gainsbury, S. M., Russell, A., Blaszczynski, A., & Hing, N. (2015). The interaction between gambling activities and modes of access : A comparison of Internet-only, land-based only, and mixed-mode gamblers. *Addictive Behaviors*, *41*, 34–40. doi: 10.1016/j.addbeh.2014.09.023

Gainsbury, S. M., Russell, A., Hing, N., Wood, R., Lubman, D. I., & Blaszczynski, A. (2014). The prevalence and determinants of problem gambling in Australia : assessing the impact of interactive gambling and new technologies. *Psychology of Addictive Behaviors : Journal of the Society of Psychologists in Addictive Behaviors*, *28*(3), 769–779. doi: 10.1037/a0036207

Garber, M. C., Nau, D. P., Erickson, S. R., Aikens, J. E., & Lawrence, J. B. (2004). The concordance of self-report with other measures of medication adherence : a summary of the literature. *Medical Care*, *42*(7), 649–652.

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

Gebauer, L., LaBrie, R. A., & Shaffer, H. J. (2010). Optimizing DSM-IV classification accuracy : A brief bio-social screen for gambling disorders among the general household population. *Can J Psychiatry*, *55*(2), 82–90.

Genolini, C., Pingault, J. B., Driss, T., Côté, S., Tremblay, R. E., Vitaro, F., ... Falissard, B. (2013). KmL3d : a non-parametric algorithm for clustering joint trajectories. *Computer Methods and Programs in Biomedicine*, *109*(1), 104–111. doi: 10.1016/j.cmpb.2012.08.016

Gerstein, D., Volberg, R. A., Toce, M. T., Harwood, H., Johnson, R. A., Buie, T., ... Engelman, L. (1999). Gambling impact and behavior study : Report to the national gambling impact study commission. *Chicago : National Opinion Research Center*.

Goodman, A. (1990). Addiction : definition and implications. *British Journal of Addiction*, *85*(11), 1403–1408. doi: 10.1111/j.1360-0443.1990.tb01620.x

Goodman, L. A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, *61*(2), 215–231.

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 : Journal of the Society of Psychologists in Addictive Behaviors*, *26*(3), 527–535. doi: 10.1037/a0028545

Griffiths, M. (2003). Internet gambling : issues, concerns, and recommendations. *Cyberpsychol Behav*, *6*(6), 557 – 568.

Griffiths, M., & Parke, J. (2002). The Social Impact of Internet Gambling. *Soc. Sci. Comput. Rev.*, *20*(3), 312–320. doi: 10.1177/089393020200300 8

Griffiths, M., Wardle, H., Orford, J., Sproston, K., & Erens, B. (2011). Internet Gambling, Health, Smoking and Alcohol Use : Findings from the 2007 British Gambling Prevalence Survey. *International Journal of Mental Health and Addiction*, *9*(1), 1–11. doi: 10.1007/s11469-009-9246-9

Haefeli, J., Lischer, S., & Schwarz, J. (2011). Early detection items and responsible gambling features for online gambling. *International Gambling Studies*, *11*(3), 273–288. doi: 10.1080/14459795.2011.604643

Hagenaars, J. A. (1988). Latent Structure Models with Direct Effects between Indicators : Local Dependence Models. *Sociological Methods & Research*, *16*(3), 379–405. doi: 10.1177/0049124188016003002

Hancock, L., Schellinck, T., & Schrans, T. (2008). Gambling and corporate social responsibility (CSR) : Re-defining industry and state roles on duty of care, host responsibility and risk management. *Policy and Society*, *27*(1), 55–68. doi: 10.1016/j.polsoc.2008.07.005

Hodgins, D. C. (2004). Using the NORC DSM Screen for Gambling Problems as an outcome measure for pathological gambling : psychometric evaluation. *Addictive Behaviors*, *29*(8), 1685–1690. doi: 10.1016/j.addbeh.2004.03.017

Hopley, A. A. B., Dempsey, K., & Nicki, R. (2012). Texas Hold'em Online Poker : A Further Examination. *International Journal of Mental Health and Addiction*, *10*(4), 563–572. doi: 10.1007/s11469-011-9353-2

Hopley, A. A. B., & Nicki, R. M. (2010). Predictive factors of excessive online poker playing. *Cyberpsychology, Behavior and Social Networking*, *13*(4), 379–385. doi: 10.1089/cyber.2009.0223

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. doi: 10.1016/j.addbeh.2015.01.011

Iancu, I., Lowengrub, K., Dembinsky, Y., Kotler, M., & Dannon, P. N. (2008). Pathological gambling : an update on neuropathophysiology and pharmacotherapy. *CNS drugs*, *22*(2), 123–138.

Inserm. (2008). *Jeux de hasard et d'argent : Contextes et addictions* (Rapport). Paris : Les éditions Inserm.

Johansson, A., Grant, J. E., Kim, S. W., Odlaug, B. L., & Götestam, K. G. (2009). Risk Factors for Problematic Gambling : A Critical Literature Review. *Journal of Gambling Studies*, *25*(1), 67–92. doi: 10.1007/s10899-008-9088-6

Johnson, E. E., Hamer, R., Nora, R. M., Tan, B., Eisenstein, N., & Engelhart, C. (1997). The Lie/Bet Questionnaire for Screening Pathological Gamblers , The Lie/Bet Questionnaire for Screening Pathological Gamblers. *Psychological Reports*, *80*(1), 83–88. doi: 10.2466/pr0.1997.80.1.83

Kairouz, S., Paradis, C., & Nadeau, L. (2012). Are online gamblers more at risk than offline gamblers ? *Cyberpsychology, Behavior and Social Networking*, *15*(3), 175–180. doi: 10.1089/cyber.2011.0260

Khazaal, Y., van Singer, M., Chatton, A., Achab, S., Zullino, D., Rothen, S., ... Thorens, G. (2014). Does Self-Selection Affect Samples' Representativeness in Online Surveys ? An Investigation in Online Video Game Research. *Journal of Medical Internet Research*, *16*(7). doi: 10.2196/jmir.2759

Laakasuo, M., Palomäki, J., & Salmela, M. (2015). Poker players with experience and skill are not "ill"–Exposing a discrepancy in measures of problem gambling. *Journal of Gambling and Commercial Gaming Research*, *1*(1).

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. doi: 10.1093/eurpub/ckn021

LaBrie, R. A., LaPlante, D. A., Nelson, S. E., Schumann, A., & Shaffer, H. J. (2007). Assessing the Playing Field : A Prospective Longitudinal Study of Internet Sports Gambling Behavior. *Journal of Gambling Studies*, *23*(3), 347–362. doi: 10.1007/s10899-007-9067-3

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

Ladouceur, R. (2004a). Gambling : The Hidden Addiction. *Can J Psychiatry*, *49*(8), 3.

Ladouceur, R. (2004b). Perceptions among pathological and nonpathological gamblers. *Addictive Behaviors*, *29*(3), 555–565. doi: 10.1016/j.addbeh.2003.08.025

Ladouceur, R., Shaffer, P., Blaszczynski, A., & Shaffer, H. J. (2017). Responsible gambling : a

synthesis of the empirical evidence. *Addiction Research & Theory*, *25*(3), 225–235. doi: 10.1080/16066359.2016.1245294

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. *Comput Hum Behav*, *25*(3), 711 – 717.

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. doi: 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. doi: 10.1016/j.chb.2008.02.015

Leisch, F., & Gruen, G. (2018). *CRAN Task View : Cluster Analysis & Finite Mixture Model.*

Lesieur, H. R., & Blume, S. B. (1987). The South Oaks Gambling Screen (SOGS) : a new instrument for the identification of pathological gamblers. *Am J Psychiatry*, *144*(9), 1184 – 1188.

Lo, Y., Mendell, N. R., & Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, *88*(3), 767–778. doi: 10.1093/biomet/88.3.767

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. doi: 10.1002/mpr.1510

Lévesque, D., Sévigny, S., Giroux, I., & Jacques, C. (2017). Gambling-Related Cognition Scale (GRCS) : Are skills-based games at a disadvantage ? *Psychology of Addictive Behaviors : Journal of the Society of Psychologists in Addictive Behaviors*, *31*(6), 647–654. doi: 10.1037/adb0000297

Magidson, J., & Vermunt, J. K. (2004). Latent class models. *The Sage handbook of quantitative methodology for the social sciences*, 175–198.

Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, *64*(2), 187–212. doi: 10.1007/BF02294535

McCormack, A., Shorter, G. W., & Griffiths, M. D. (2013). An examination of participation in online gambling activities and the relationship with problem gambling. *Journal of Behavioral Addictions*, *2*(1), 31–41. doi: 10.1556/JBA.2.2013.1.5

McCutcheon, A. L. (1987). *Latent Class Analysis*. SAGE. (Google-Books-ID : dCbJ6NcH4mAC)

McLachlan, G., & Peel, D. (2000). *Finite mixture models, willey series in probability and statistics*. John Wiley & Sons, New York.

Merkouris, S., Thomas, S., Browning, C., & Dowling, N. (2016). Predictors of outcomes of psychological treatments for disordered gambling : A systematic review. *Clinical Psychology*

Review, *48*, 7–31. doi: 10.1016/j.cpr.2016.06.004

Miller, H. (2015). *Risk factors for problem gambling : Environmental, geographic, social, cultural, demographic, socio-economic, family and household* (Rapport). Australia : Victorian Responsible Gambling Foundation.

Mitrovic, D. V., & Brown, J. (2009). Poker mania and problem gambling : a study of distorted cognitions, motivation and alexithymia. *Journal of Gambling Studies*, *25*(4), 489–502. doi: 10.1007/s10899-009-9140-1

Moreau, A., Chabrol, H., & Chauchard, E. (2016). Psychopathology of Online Poker Players : Review of Literature. *Journal of Behavioral Addictions*, *5*(2), 155–168. doi: 10.1556/2006.5.2016.035

Moreau, A., Delieuvin, J., Chabrol, H., & Chauchard, E. (2017). Online Poker Tilt Scale (OPTS) : creation and validation of a tilt assessment in a French population. *International Gambling Studies*, *17*(2), 205–218. doi: 10.1080/14459795.2017.1321680

Moreau, A., Delieuvin, J., Chauchard, E., & Chabrol, H. (2015). Le "tilt" au poker en ligne. Un comportement pathologique transitoire ? *Alcoologie et Addictologie*, *37*(3), 245–251.

Muthen, L. K., & Muthen, B. O. (2017). *Mplus User's Guide. Eighth Edition* (Rapport). Los Angeles, CA.

Muthén, B., & Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, *55*(2), 463–469.

Nagin, D. S., & Land, K. C. (1993). Age, Criminal Careers, and Population Heterogeneity : Specification and Estimation of a Nonparametric, Mixed Poisson Model. *Criminology*, *31*(3), 327–362. doi: 10.1111/j.1745-9125.1993.tb01133.x

Nelson, S. E., LaPlante, D. A., Peller, A. J., Schumann, A., LaBrie, R. A., & Shaffer, H. J. (2008). Real limits in the virtual world : self-limiting behavior of internet gamblers. *J Gambl Stud*, *24*(4), 463 – 477.

Nunnally Jr., J. C. (1970). *Introduction to psychological measurement.* New York, NY, US : McGraw-Hill.

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling : A Monte Carlo simulation study. *Structural equation modeling*, *14*(4), 535–569.

Oberski, D. L. (2016). Beyond the number of classes : separating substantive from non-substantive dependence in latent class analysis. *Advances in Data Analysis and Classification*, *10*(2), 171–182. doi: 10.1007/s11634-015-0211-0

Olievenstein, C. (1983). *La drogue ou la vie.* Paris : Éditions Robert Laffont.

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. doi: 10.1007/s11749-014-0387-1

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. doi: 10.1080/14459795.2017.1378362

Perrot, B., Bataille, E., & Hardouin, J.-B. (2018). validscale : A command to validate measurement scales. *Stata Journal*, *18*(1), 29–50.

Philander, K. S. (2014). Identifying high-risk online gamblers : a comparison of data mining procedures. *International Gambling Studies*, *14*(1), 53–63. doi: 10.1080/14459795.2013.841721

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. doi: 10.1080/14459795.2014.893585

Problem gambling is a public health concern. (2017). *The Lancet*, *390*(10098), 913. doi: 10.1016/S0140-6736(17)32333-4

Proust-Lima, C., Amieva, H., & Jacqmin-Gadda, H. (2013). Analysis of multivariate mixed longitudinal data : a flexible latent process approach. *The British Journal of Mathematical and Statistical Psychology*, *66*(3), 470–487. doi: 10.1111/bmsp.12000

Proust-Lima, C., Philipps, V., & Liquet, B. (2017). Estimation of Extended Mixed Models Using Latent Classes and Latent Processes : The R Package lcmm. *Journal of Statistical Software ; Vol 1, Issue 2 (2017).*

Raylu, N., & Oei, T. P. (2004). The Gambling Related Cognitions Scale (GRCS) : Development, confirmatory factor validation and psychometric properties. *Addiction*, *99*(6), 757–769.

Rundle-Thiele, S. (2009). Bridging the gap between claimed and actual behaviour : The role of observational research. *Qualitative Market Research : An International Journal*, *12*(3), 295–306. doi: 10.1108/13522750910963818

Science has a gambling problem. (2018). *Nature*, *553*(7689), 379. doi: 10.1038/d41586-018-01051-z

Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). mclust 5 : Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *The R journal*, *8*(1), 289–317.

Shaffer, H. J., Peller, A. J., LaPlante, D. A., Nelson, S. E., & LaBrie, R. A. (2010). Toward a paradigm shift in Internet gambling research : From opinion and self-report to actual behavior. *Addiction Research & Theory*, *18*(3), 270–283. doi: 10.3109/16066350902777974

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

Shead, N. W., Hodgins, D. C., & Scharf, D. (2008). Differences between Poker Players and Non-Poker-Playing Gamblers. *International Gambling Studies*, *8*(2), 167–178. doi: 10.1080/14459790802139991

Stinchfield, R., McCready, J., & Turner, N. (2012). A comprehensive review of problem gambling

screens and scales for online self-assessment. *Toronto : Ontario Problem Gambling Research Centre*.

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

Thomas, S., Jackson, A., & Blaszczynski, A. (2003). Measuring problem gambling : Evaluation of the Victorian gambling screen. *Report for the Gambling Research Panel (GRP Report No. 4)*.

Thomas, S., Merkouris, S., Dowling, N., Radermacher, H., Jackson, A., Misso, M., & Anderson, C. (2011). *Guideline for screening, assessment and treatment in problem gambling*. Monash University.

Toce-Gerstein, M., Gerstein, D. R., & Volberg, R. A. (2009). The NODS–CLiP : A Rapid Screen for Adult Pathological and Problem Gambling. *Journal of Gambling Studies*, *25*(4), 541. doi: 10.1007/s10899-009-9135-y

Tovar, M. L., Costes, J. M., & Eroukmanoff, V. (2013). Les jeux d'argent et de hasard sur internet en France en 2012. *Tendances*, *85*, 1 – 6.

Valleur, M., & Matysiak, J.-C. (2006). *Les addictions : panorama clinique, modèles explicatifs, débat social et prise en charge*. A. Colin.

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*.

Verbeke, G., & Lesaffre, E. (1996). A Linear Mixed-Effects Model with Heterogeneity in the Random-Effects Population. *Journal of the American Statistical Association*, *91*(433), 217–221. doi: 10.1080/01621459.1996.10476679

Vermunt, J. K. (2003). Multilevel Latent Class Models. *Sociological Methodology*, *33*(1), 213–239. doi: 10.1111/j.0081-1750.2003.t01-1-00131.x

Vermunt, J. K. (2004). An EM algorithm for the estimation of parametric and nonparametric hierarchical nonlinear models. *Statistica Neerlandica*, *58*(2), 220–233. doi: 10.1046/j.0039-0402.2003.00257.x

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

Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. Dans *Applied Latent Class Analysis* (Vol. 11, pp. 89–106).

Vermunt, J. K., & Magidson, J. (2015). *Technical Guide for Latent GOLD 5.1 : Basic, Advanced, and Syntax*. Belmont, MA : Statistical Innovations Inc.

Wardle, H., Sproston, K., Orford, J., Erens, B., Griffiths, M., Constantine, R., & Pigott, S. (2007). *British Gambling Prevalence Survey 2007* (Rapport).

Weinstock, J., Massura, C. E., & Petry, N. M. (2013). Professional and pathological gamblers : similarities and differences. *Journal of Gambling Studies*, *29*(2), 205–216. doi: 10.1007/s10899-012-9308-y

Wood, R. T., Griffiths, M., & Parke, J. (2007). Acquisition, Development, and Maintenance of Online Poker Playing in a Student Sample. *CyberPsychology & Behavior*, *10*(3), 354–361. doi: 10.1089/cpb.2006.9944

Wood, R. T., & Williams, R. J. (2009). *Internet gambling : Prevalence, patterns, problems, and policy options* (Rapport). Final Report prepared for the Ontario Problem Gambling Research Centre ; Guelph, Ontario.

Xuan, Z., & Shaffer, H. (2009). How Do Gamblers End Gambling : Longitudinal Analysis of Internet Gambling Behaviors Prior to Account Closure Due to Gambling Related Problems. *Journal of Gambling Studies*, *25*(2), 239–252. doi: 10.1007/s10899-009-9118-z

Yani-de Soriano, M., Javed, U., & Yousafzai, S. (2012). Can an Industry Be Socially Responsible If Its Products Harm Consumers ? The Case of Online Gambling. *Journal of Business Ethics*, *110*(4), 481–497. doi: 10.1007/s10551-012-1495-z

Young, M., Stevens, M., & Morris, M. (2008). Problem Gambling within the Non-Indigenous Population of the Northern Territory of Australia : A Multivariate Analysis of Risk Factors. *International Gambling Studies*, *8*(1), 77–93. doi: 10.1080/14459790701870571

Annexes

Annexe A

Indice Canadien du Jeu Excessif (Ferris & Wynne, 2001)

Au cours des 12 derniers mois :

- Avez-vous misé plus d'argent que vous pouviez vous permettre de perdre ?
- Avez-vous besoin de miser de plus en plus d'argent pour avoir la même excitation ?
- Avez-vous rejoué une autre journée pour récupérer l'argent que vous aviez perdu en jouant ?
- Avez-vous vendu ou emprunté quelque chose pour obtenir de l'argent pour jouer ?
- Avez-vous déjà senti que vous aviez peut-être un problème avec le jeu ?
- Le jeu a-t-il causé chez vous des problèmes de santé, y compris du stress ou de l'angoisse ?
- Des personnes ont-elles critiqué vos habitudes de jeu ou dit que vous aviez un problème avec le jeu ?
- Vos habitudes de jeu ont-elles causé des difficultés financières à vous ou à votre entourage ?
- Vous êtes-vous déjà senti coupable de vos habitudes de jeu ou de ce qui arrive quand vous jouez ?

Réponses et cotation (par question) :

- Jamais (0)
- Parfois (1)
- La plupart du temps (2)
- Presque toujours (3)

Annexe B

Article : Harmonizing Screening for Gambling Problems in Epidemiological Surveys – Development of the Rapid Screener for Problem Gambling (RSPG)

FULL-LENGTH REPORT

Epub ahead of print: June 27, 2016
Journal of Behavioral Addictions 5(2), pp. 239–250 (2016)
DOI: 10.1556/2006.5.2016.030

Harmonizing Screening for Gambling Problems in Epidemiological Surveys – Development of the Rapid Screener for Problem Gambling (RSPG)

GAËLLE CHALLET-BOUJU[1,2]*, BASTIEN PERROT[2,3], LUCIA ROMO[4,5,6], MARC VALLEUR[7], DAVID MAGALON[8], MÉLINA FATSÉAS[9], ISABELLE CHÉREAU-BOUDET[10], AMANDINE LUQUIENS[11], JEU GROUP, MARIE GRALL-BRONNEC[1,2] and JEAN-BENOIT HARDOUIN[2,3]

[1]CHU Nantes, Clinical Investigation Unit BALANCED "BehaviorAL AddictioNs and ComplEx mood Disorders," Department of Addictology and Psychiatry, France
[2]EA 4275 SPHERE "bioStatistics, Pharmacoepidemiology and Human sciEnces Research tEam," Faculties of Medicine and Pharmaceutical Sciences, University of Nantes, France
[3]Unit of Methodology and Biostatistics, University Hospital of Nantes, France
[4]EA 4430 CLIPSYD "CLInique PSYchanalyse Développement," University of Paris Ouest Nanterre La Défense, France
[5]Louis Mourier Hospital of Colombes, Assistance Publique – Hôpitaux de Paris (APHP), France
[6]Psychotherapies Unit, Sainte-Anne Hospital – Psychiatry and Neurosciences, Paris, France
[7]Marmottan Medical Center, GPS Perray-Vaucluse, Paris, France
[8]Department of Adult Psychiatry, Sainte-Marguerite University Hospital of Marseille, France
[9]Psychiatry Laboratory, Sanpsy CNRS USR 3413, University of Bordeaux and Charles Perrens Hospital, Bordeaux, France
[10]Psychiatry Department, University Hospital of Clermont-Ferrand, France
[11]Psychiatry and Addictology Department, Paul Brousse University Hospital of Villejuif, Assistance Publique – Hôpitaux de Paris (APHP), France

(Received: July 27, 2015; accepted: March 22, 2016)

Background and aims: The aim of this study was to test the screening properties of several combinations of items from gambling scales, in order to harmonize screening of gambling problems in epidemiological surveys. The objective was to propose two brief screening tools (three items or less) for a use in interviews and self-administered questionnaires. *Methods:* We tested the screening properties of combinations of items from several gambling scales, in a sample of 425 gamblers (301 non-problem gamblers and 124 disordered gamblers). Items tested included interview-based items (Pathological Gambling section of the DSM-IV, lifetime history of problem gambling, monthly expenses in gambling, and abstinence of 1 month or more) and self-report items (South Oaks Gambling Screen, Gambling Attitudes, and Beliefs Survey). The gold standard used was the diagnosis of a gambling disorder according to the DSM-5. *Results:* Two versions of the Rapid Screener for Problem Gambling (RSPG) were developed: the RSPG-Interview (RSPG-I), being composed of two interview items (increasing bets and loss of control), and the RSPG-Self-Assessment (RSPG-SA), being composed of three self-report items (chasing, guiltiness, and perceived inability to stop). *Discussion and conclusions:* We recommend using the RSPG-SA/I for screening problem gambling in epidemiological surveys, with the version adapted for each purpose (RSPG-I for interview-based surveys and RSPG-SA for self-administered surveys). This first triage of potential problem gamblers must be supplemented by further assessment, as it may overestimate the proportion of problem gamblers. However, a first triage has the great advantage of saving time and energy in large-scale screening for problem gambling.

Keywords: screening, problem gambling, epidemiological surveys, general population, RSPG

INTRODUCTION

Previously known as "Pathological Gambling," Gambling Disorder has recently been included in the "substance-related and addictive disorders" section of the DSM-5 (American Psychiatric Association, 2013). The prevalence of lifetime pathological gambling has been estimated at around 0.4–1.0% (American Psychiatric Association, 2013), depending on the country, the sample assessed, and particularly the assessment tool used. Indeed, one limitation when working on general population-based samples (like those used for prevalence surveys) is that long structured diagnostic interviews [like the Structured Clinical Interview for Pathological Gambling (SCI-PG) (Grant, Steinberg,

* Corresponding author: Gaëlle Challet-Bouju; Unité d'Investigation Clinique "Addictions comportementales et Troubles Complexes de l'Humeur," Institut Fédératif des Addictions Comportementales (IFAC), CHU de Nantes–Hôpital Saint Jacques, Bât. Louis Philippe, 85 rue de Saint Jacques, 44093 Nantes Cedex 1, France; Phone: +33 2 40 84 76 20; Fax: +33 2 40 84 61 18; E-mail: gaelle.bouju@chu-nantes.fr

Members of the JEU Group: Marie Grall-Bronnec, Gaëlle Challet-Bouju, Jean-Luc Vénisse, Lucia Romo, Cindy Legauffre, Caroline Dubertret, Irène Codina, Marc Valleur, Marc Auriacombe, Mélina Fatséas, Jean-Marc Alexandre, Pierre-Michel Llorca, Isabelle Chéreau-Boudet, Christophe Lançon, David Magalon, Michel Reynaud et Amandine Luquiens.

© 2016 The Author(s)

Kim, Rounsaville, & Potenza, 2004) or the Diagnostic Interview for Gambling Severity (DIGS) (Winters, Specker, & Stinchfield, 2002)] are not feasible on a large scale, because they need significant and costly human resources. Thus, several self-report tools have been developed to enable rapid identification of gambling problems, which can be classified into three categories according to the Problem Gambling Research and Treatment Centre (PGRTC, 2011). Examples of long (>13 items) questionnaires are the South Oaks Gambling Screen (SOGS) (Lesieur & Blume, 1987), the National Opinion Research Center DSM Screen for Gambling Problems (NODS) (Gerstein et al., 1999; Hodgins, 2004), and the Victorian Gambling Screen (VGS) (Tolchard & Battersby, 2010). An example of medium-sized (4–12 items) questionnaires is the Canadian Problem Gambling Inventory (CPGI) (Ferris & Wynne, 2001). Finally, examples of brief (1–3 items) screening tools are the one-item Screen for Problem Gambling (one-item SPG) (Thomas, Piterman, & Jackson, 2008), the Lie/Bet Questionnaire (Johnson et al., 1997), the Brief Biosocial Gambling Screen (BBGS) (Gebauer, LaBrie, & Shaffer, 2010), and the NODS-CLiP ("Control, Lying, and Preoccupation" version of the NODS – Toce-Gerstein, Gerstein, & Volberg, 2009). Because of their short duration, brief screening tools provide a good alternative for systematic screening of large samples of people who may have an elevated risk of gambling-related problems. They must be supplemented by further assessment in case of a positive screening (PGRTC, 2011) and should only be used as a first triage of people to be interviewed.

According to Stinchfield, McCready, and Turner (2012), the majority of the brief problem gambling screens have not been rigorously tested beyond their development. It is currently unknown what screen to be chosen for what kind of population, and whether a cross-use in clinical and community-based samples is possible without decreasing the psychometric properties. It is also still unknown if such tools are usable interchangeably in interview-based (like telephone-based surveys) or self-assessment-based (like internet-based surveys) epidemiological surveys, outside the context of their initial development.

For example, the Lie/Bet Questionnaire (Johnson et al., 1997) is often used in general population for a very quick screening of problem gambling, even though Stinchfield et al. (2012) rightly noted that this test is not as accurate in screening for problem gambling among the general population (because the authors did not take the low prevalence of problem gambling into account in the validation studies). Moreover, they emphasized that the Lie/Bet Questionnaire displays several limitations, especially the use of a lifetime timeframe (which is likely to increase false positive rate) and the lack of empirical evidence of classification accuracy from investigators other than the developers. Moreover, no recommendation is done about a use as a self-assessment questionnaire or an interview.

The NODS-CLiP is certainly the mostly used brief screening tool for PG. But it also displays some limitations, for example, the lifetime timeframe, the lack of independence of the gold standard used for the validation study (the full NODS), and the lower screening performance in a clinical sample (Stinchfield et al., 2012).

The BBGS is the more recently developed brief tool. It has the advantage of including a current timeframe (past 12 months). There is again a lack of independence of the gold standard used for the validation study (the full AUDADIS from which the BBGS is derived) (Stinchfield et al., 2012). Due to its recent development, the BBGS also requires further research to cross-validate its classification accuracy estimates (Stinchfield et al., 2012), because of the limited date available (PGRTC, 2011).

Finally, the one-item SPG is one of the least used brief screening tools. Critics about the one-item SPG are mainly about its important brevity (just one item), which yields high error rates, particularly false negative rates (Stinchfield et al., 2012).

The aim of our work is to perform a systematic search of the best combinations of items from several gambling scales, to develop two brief screening tools (three items or less) being suitable for interview-based epidemiological surveys on the one hand and self-rating-based epidemiological surveys on the other hand. Indeed, self-assessment and interviews are not strictly equivalent and the tools used must be adapted for each need. The objective is to harmonize screening of gambling problems in epidemiological surveys. Another objective is to compare the combinations tested to equivalents of the four major brief screening scales previously validated and widely used (one-item SPG, Lie/Bet, BBGS, and NODS-CLiP) within a single sample, with good representation of problem gamblers. Indeed, to our knowledge, those tools have so far neither been compared to each other in the same dataset, nor tested in samples combining problem gamblers and non-problem gamblers (NPG).

METHODS

Participants

This analysis was performed using baseline data from the JEU cohort study, collected between 2009 and 2011 (Challet-Bouju et al., 2014). The JEU study is a 5-year longitudinal cohort study performed at national level. The JEU cohort consists of 628 gamblers, divided into three groups, depending on whether they are problem or non-problem gamblers and on whether they had sought treatment or not. The main objective of this study is to identify the determinants of key state changes in the gambling practice. The proposed analysis is thus a secondary data analysis.

Initially based on the Pathological Gambling section of the DSM-IV, the diagnosis of a gambling disorder was reassessed for this analysis according to the current DSM-5 section for Gambling Disorders (i.e., removal of criterion 8 on illegal acts and use of a threshold of 4). With this re-categorization, the sample was composed of 301 NPG, 124 problem gamblers without treatment (PGWT), and 186 problem gamblers seeking treatment (PGST) (17 more gamblers were initially in the PGST group, but were reclassified as NPG with the DSM-5 classification; we decided to exclude them *a posteriori* from the NPG group because they had sought treatment).

NPG and PGWT were recruited in various gambling places and via the press, in order to cover the broadest

possible range of gambling activities. PGSTs were recruited in seven care centers, among patients who started treatment less than 6 months before. Only participants who reported gambling on at least one occasion during the previous year and who were between 18 and 65 years old were included in the study. For more information about the JEU cohort study, refer to the study protocol (Challet-Bouju et al., 2014).

Measures

Among other assessments, participants were interviewed throughout a past year – interview based on the 10 diagnostic criteria for Pathological Gambling in the DSM-IV (American Psychiatric Association, 2000), an interview designed to explore their gambling history and habits (lifetime subjective history of problem gambling, monthly expenses in gambling, and abstinence of 1 month or more), and two self-report questionnaires: the SOGS (Lejoyeux, 1999; Lesieur & Blume, 1987) and the Gambling Attitudes and Beliefs Scale (GABS) (Bouju et al., 2014; Breen & Zuckerman, 1999). The SOGS is a well-known measure of problem gambling severity, and the GABS measures cognitive biases, irrational beliefs, and positively valued attitudes to gambling (Breen & Zuckerman, 1999).

The formulations of the interview-based items are given in Table 1 [for the formulations of items from the SOGS and the GABS, see the validation papers of these scales (Breen & Zuckerman, 1999; Lesieur & Blume, 1987)]. The DSM-5 status was used as a gold standard to perform the screening properties analysis. All the scales were administered in French language.

For the secondary objective, we focused particularly on combinations which could correspond to equivalents of the four major brief screening scales (one-item SPG, Lie/Bet, BBGS, and NODS-CLiP). Indeed, even if these combinations are not strictly formulated in the same way as the original scales, we presume that they represent good equivalents that could be used for the comparison. The formulations of the originals and equivalents of the four brief screening tools are given in Table 2.

Table 1. Formulations of the interview-based items

Gambling history and habits (lifetime history of problem gambling): "Has your gambling practice already become a problem for you?"
Gambling history and habits (monthly expenses in gambling): "How much money do you bet on average per month in gambling?"
Gambling history and habits (abstinence of 1 month or more): "Have you ever fully stopped gambling for at least one month?"
Item 1 from the DSM-IV (preoccupation): "During the past 12 months, have you been constantly preoccupied with gambling (e.g., preoccupied with reliving past gambling experiences, planning the next venture, or thinking of ways to get money with which to gamble)?"
Item 2 from the DSM-IV (increasing bets): "During the past 12 months, have you needed to gamble with increasing amounts of money in order to achieve the desired excitement?"
Item 3 from the DSM-IV (loss of control): "During the past 12 months, have you made repeated unsuccessful efforts to control, cut back, or stop gambling?"
Item 4 from the DSM-IV (withdrawal): "During the past 12 months, have you been restless or irritable when attempting to cut down or stop gambling?"
Item 5 from the DSM-IV (escapism): "During the past 12 months, did you gamble as a way of escaping from the hardships of life or of relieving an undesirable mood (e.g., feelings of helplessness, guilt, anxiety, depression)?"
Item 6 from the DSM-IV (chasing): "During the past 12 months, after having lost money in gambling, did you often return to gambling another day in order to get even, to recover your losses?"
Item 7 from the DSM-IV (lying): "During the past 12 months, have you lied to family member, therapist, or others to conceal the extent of involvement with gambling?"
Item 8 from the DSM-IV (illegal acts): "During the past 12 months, have you committed illegal acts, such as forgery, fraud, theft, or embezzlement, in order to finance gambling?"
Item 9 from the DSM-IV (jeopardizing of other activities): "During the past 12 months, have you jeopardized or lost a significant relationship, job, or educational or career opportunity because of gambling?"
Item 10 from the DSM-IV (borrowing money): "During the past 12 months, have you relied on others to provide money to relieve a desperate financial situation caused by gambling?"

Statistical analysis

For the interview-based tool, we tested the screening properties of combinations of the 10 items from the Pathological Gambling section of the DSM-IV and the three questions from the gambling history and habits (lifetime history of problem gambling, monthly expenses in gambling, and abstinence of 1 month or more). For the self-report-based tool, we tested the screening properties of combinations of 25 items from the SOGS (items 2–16 k, excluding question 12 and including non-scoring items) and the 35 items from the GABS. Question 2 of the SOGS was rephrased in order to obtain the value of the maximum amount of money wagered in a day (instead of amounts grouped in classes). In order to obtain a binary response to this question, we determined a threshold of €180 to have a "yes" answer, calculated to maximize the Youden Index [(YI) which is a statistic that measures the performance of a diagnostic test, independently of the disorder's prevalence] (Youden, 1950). The same strategy was used for the interview question about the monthly expenses in gambling, with a threshold of €125. GABS responses (with four response modalities from "strongly agree" to "strongly disagree") were redefined as "yes" answers when "strongly agree" or "agree" were chosen for reversed items, and vice versa for direct items.

Table 2. Formulations of the original and equivalents four major brief screening tools tested

One-item SPG (Rockloff, Ehrich, Themessl-Huber, & Evans, 2011)
Original item:
"In the past 12 months, have you ever had an issue with your gambling?"

Item used in the JEU cohort (= used as an equivalent of the one-item SPG):
"Do you feel you have ever had a problem with betting or money gambling in the past 12 months?" *(=item 6 from the SOGS)*

Lie/Bet Questionnaire (Johnson et al., 1997)
Original items:
Lie: "Have you ever had to lie to people important to you about how much you gambled?"
Bet: "Have you ever felt the need to bet more and more money?"

Items used in the JEU cohort (= used as an equivalent of the Lie/Bet Questionnaire):
Lie: "During the past 12 months, have you lied to family members, therapist, or others to conceal the extent of involvement with gambling?" *(=item 7 from the DSM-IV)*
Bet: "During the past 12 months, have you needed to gamble with increasing amounts of money in order to achieve the desired excitement?" *(=item 2 from the DSM-IV)*

BBGS (Gebauer et al., 2010)
Original items:
Withdrawal: "During the past 12 months, have you become restless, irritable, or anxious when trying to stop and/or cut down on gambling?"
Lying: "During the past 12 months, have you tried to keep your family or friends from knowing how much you gambled?"
Borrowing money: "During the past 12 months, did you have such financial trouble as a result of gambling that you had to get help with living expenses from family, friends, or welfare?"

Items used in the JEU cohort (= used as an equivalent of the BBGS):
Withdrawal: "During the past 12 months, have you been restless or irritable when attempting to cut down or stop gambling?" *(=item 4 from the DSM-IV)*
Lying: "During the past 12 months, have you lied to family members, therapist, or others to conceal the extent of involvement with gambling?" *(=item 7 from the DSM-IV)*
Borrowing money: "During the past 12 months, have you relied on others to provide money to relieve a desperate financial situation caused by gambling?" *(=item 10 from the DSM-IV)*

NODS-CLiP (Toce-Gerstein et al., 2009)
Original items:
Preoccupation: "Have there ever been periods lasting 2 weeks or longer when you spent a lot of time thinking about your gambling experiences or planning out future gambling ventures or bets?"
Loss of control: "Have you ever tried to stop, cut down, or control your gambling?"
Lying: "Have you ever lied to family members, friends, or others about how much you gamble or how much money you lost on gambling?"

Items used in the JEU cohort (= used as an equivalent of the NODS-CLiP):
Preoccupation: "During the past 12 months, have you been constantly preoccupied with gambling (e.g., preoccupied with reliving past gambling experiences, planning the next venture, or thinking of ways to get money with which to gamble)?" *(=item 1 from the DSM-IV)*
Loss of control: "During the past 12 months, have you made repeated unsuccessful efforts to control, cut back or stop gambling?" *(=item 3 from the DSM-IV)*
Lying: "During the past 12 months, have you ever lied to family members, therapist, or others to conceal the extent of involvement with gambling?" *(=item 7 from the DSM-IV)*

All possible combinations of one to three items ($n = 377$ for interview combinations and $n = 34{,}279$ for self-report combinations) were tested. A combination was considered as having a positive screening when there was at least one "yes" answer. Screening properties (Sensitivity: Se, Specificity: Sp, Incorrect Classification Rate: ICR, Youden Index: YI, and Positive and Negative Predictive Values: PPV and NPV) were estimated in the JEU sample excluding PGSTs. We excluded PGSTs because the final brief screening tools are not intended to be used clinically for problem gamblers in treatment, so that their inclusion would not have been relevant.

In addition to providing the screening properties in our sample without the PGSTs, we then re-computed PPV, NPV, and ICR to take into account the French prevalence of problem gambling in the population of past-year gamblers, that is, 2.76% (Costes et al., 2011). The prevalence of problem gambling in past-year gamblers (Pr) was calculated as follows: prevalence of active gamblers in the general population (12.2%) × prevalence of problem gambling in active gamblers (10.8%)/prevalence of past-year gamblers in the general population (47.8%) (Costes et al., 2011). The adjusted PPV was calculated as follows: $Se \times Pr/[Se \times Pr + (1 - Sp)(1 - Pr)]$, the adjusted NPV as follows: $Sp(1 - Pr)/[Sp(1 - Pr) + (1 - Se)Pr]$, and the adjusted ICR as follows: $(1 - Sp)(1 - Pr) + (1 - Se)Pr$, where $Pr = 0.0276$. The objective, in adjusting the proportion of problem gamblers according to national prevalence, was to obtain a more

realistic sample (because the sample used in the JEU cohort study was constituted based on an approximate equality of size between problem and non-problem gamblers, it overestimates the prevalence of problem gambling).

As stated above, the aim of brief screening tools is to enable rapid identification of potential problem gamblers, in order to reduce the number of people to be interviewed. The objective is therefore to miss as few potential problem gamblers as possible. Consequently, we chose to retain only the combinations of items with the highest sensitivity for consideration (we set the threshold at 0.95). Of these, the final choice of the best combination for each modality (interview or self-report) and within each number of items was the one with the highest YI, which is a statistic that combines both Se and Sp, and thus gives a good compromise between true positive and true negative rates. Within the best retained combinations of each modality, we chose to favor the briefer ones if they had equivalent properties, in order to have the shortest duration. As argued by Volberg, Munck, and Petry (2011), although sensitivity is the key performance characteristic of a screening tool, diagnostic efficiency is also very important to determine the tool's accuracy. Thus, the equivalence of properties was determined as a difference in Se, Sp, or ICR of less than 5%.

Ethics

Participants were informed about the research and gave their written informed consent prior to their inclusion in the study. This study was approved by the French Research Ethics Committee (CPP) on January 8, 2009.

RESULTS

Combinations of interview-based items

There were 194 interview combinations with sensitivity greater or equal to 0.95, 15 of them being composed of two items, and 179 of three items (none with just one item). Among them, the top 10 combinations (regarding their YI) are presented in Table 3 (top 5 for 2-item combinations and top 5 for 3-item combinations), with their relative screening properties. We can notice that the combinations which were considered as equivalents to three of the four major brief screening tools (Lie/Bet, BBGS, and NODS-CLiP) were not in the top 10. The equivalent of the NODS-CLiP was the only one to be in the 194 combinations with sensitivity greater or equal to 0.95 (in the 49th position of 3-item combinations).

The best combinations were composed of items 2 (increasing bets) and 3 (loss of control) from the DSM for 2-item combinations, and items 1 (preoccupation), 4 (withdrawal), and 9 (jeopardizing of other activities) from the DSM for 3-item combinations.

Since both combinations exhibited equivalent properties (difference for Se of 3% and Sp and ICR of 2%), we chose to retain the briefer combination (i.e., DSM2–DSM3 combination) as the best one for interview-based brief screening tool. We named it the Rapid Screener for Problem Gambling – Interview (RSPG-I).

Combinations of self-assessment items

There were 7,465 self-assessment combinations with sensitivity greater or equal to 0.95, 100 of them being composed of two items and 7,365 of three items (none with just one item). Among them, the top 10 combinations (regarding their YI) are presented in Table 4 (top 5 for 2-item combinations and top 5 for 3-item combinations), with their relative screening properties. We can notice that the combination which was considered as equivalent to one of the major brief screening tools (one-item) was not in the top 10, and even not in the high-sensitivity combinations.

The best combinations were composed of item 6 from the SOGS (past 12 months history of PG) and item 5 from the GABS (losing track of time) for 2-item combinations, and items 4 (chasing), 9 (guiltiness), and 10 (perceived inability to stop) from the SOGS for 3-item combinations.

Since the two combinations exhibited very different properties (difference for Se of 1%, Sp of 13%, and ICR of 9%), we chose to retain the longer combination (that is SOGS4–SOGS9–SOGS10) as the best one for self-assessment brief screening tool. We named it the Rapid Screener for Problem Gambling – Self-Assessment (RSPG-SA).

Classification accuracy

Table 5 shows the concordance between DSM-5 diagnosis and RSPG-I and RSPG-SA classifications. The results show that the RSPG-I and the RSPG-SA captured the vast majority (95.2% and 95.0%, respectively) of participants with a gambling disorder according to the DSM-5. It means that less than 5% of disordered gamblers were not covered by the RSPG. Both RSPG versions captured a significant proportion of gamblers without a gambling disorder (21.9% for RSPG-I and 36.3% for RSPG-SA), which reflects the low specificity of the two questionnaires, particularly RSPG-SA. As specificity performance was not the priority for the selection of best combinations, this is not surprising. The objective was to miss as few disordered gamblers as possible. Concerning the severity assessed by the DSM-5 among the participants with a gambling disorder, all the severe disordered gamblers were captured by both questionnaires and almost all the moderate disordered gamblers were captured by RSPG-I (97.7%) and RSPG-SA (97.6%). For mild disordered gamblers, the capture rate was lower for both questionnaires, but still acceptable (91.2% and 91.1%).

DISCUSSION AND CONCLUSIONS

The objective of this work was to develop two brief screening tools (three items or less) being suitable for reducing the duration of problem gambling screening in interview-based and self-rating-based epidemiological surveys. The analyses have enabled us to propose two very brief screening tools (two or three items) for problem gambling: the RSPG-I and the RSPG-SA. The final RSPG-I and RSPG-SA are given in English language in Table 6 and in French language in Table 7 (note that the French version was the only one to be tested in this work, and that the English version corresponds to a translation which has not been validated).

Table 3. Screening properties of the top 10 best combinations of interview-based items, and comparison with three equivalents of Brief Screening Tools (Lie/Bet, BBGS, and NODS-CLiP)

	Within the sample without PGSTs ($n = 425$)						Re-computation with national prevalence		
	YI	Se	Sp	PPV	NPV	ICR	aPPV	aNPV	aICR
Top 10 combinations of interview-based items									
2-Item combinations									
DSM2 (increasing bets) and DSM3 (loss of control)	*0.732*	*0.952*	*0.781*	*0.642*	*0.975*	*16.9%*	*0.110*	*0.998*	*21.5%*
DSM1 (preoccupation) and DSM7 (lying)	0.724	0.960	0.764	0.627	0.979	17.9%	0.104	0.999	23.0%
DSM3 (loss of control) and lifetime history of PG	0.719	0.968	0.751	0.616	0.983	18.6%	0.099	0.999	24.3%
DSM1 (preoccupation) and DSM3 (loss of control)	0.711	0.960	0.751	0.614	0.978	18.8%	0.099	0.998	24.3%
DSM3 (loss of control) and DSM7 (lying)	0.706	0.952	0.754	0.615	0.974	18.8%	0.099	0.998	24.0%
3-Item combinations									
DSM1 (preoccupation), DSM4 (withdrawal), and DSM9 (jeopardizing of other activities)	0.773	0.976	0.797	0.665	0.988	15.1%	0.120	0.999	19.8%
DSM2 (increasing bets), DSM3 (loss of control), and DSM9 (jeopardizing of other activities)	0.745	0.984	0.761	0.629	0.991	17.4%	0.105	0.999	23.3%
DSM2 (increasing bets), DSM3 (loss of control), and DSM10 (borrowing money)	0.742	0.968	0.774	0.639	0.983	16.9%	0.108	0.999	22.1%
DSM2 (increasing bets), DSM3 (loss of control), and DSM8 (illegal acts)	0.742	0.968	0.774	0.639	0.983	16.9%	0.108	0.999	22.1%
DSM2 (increasing bets), DSM7 (lying), and DSM10 (borrowing money)	0.739	0.952	0.787	0.649	0.975	16.5%	0.113	0.998	20.8%
Equivalents of brief screening tools[a]									
DSM2 (increasing bets) and DSM7 (lying) = equivalent of the Lie/Bet Questionnaire (Johnson et al., 1997)	0.718	0.927	0.791	0.646	0.964	16.9%	0.112	0.997	20.6%
DSM1 (preoccupation), DSM3 (loss of control), and DSM7 (lying) = equivalent to the NODS-CLiP (Toce-Gerstein et al., 2009)	0.666	0.992	0.674	0.557	0.995	23.3%	0.080	1.000	31.7%
DSM4 (withdrawal), DSM7 (lying), and DSM10 (borrowing money) = equivalent of the BBGS (Gebauer et al., 2010)	0.731	0.927	0.804	0.661	0.964	16.0%	0.118	0.997	19.3%

Notes. Only combinations with the highest sensitivity (Se ≥ 0.95) were first selected, and the obtained list of combinations was then graded according to the highest Youden Index (YI). The retained combination is emphasized in italic font.
For all combinations, the screening of a gambling problem was based on at least one "yes" answer.
DSM#: DSM criterion number in the order of the DSM-IV.
YI: Youden Index = Sensitivity + Specificity − 1; performance of a diagnostic test, independently of the disorder's prevalence.
Se: Sensitivity; probability of screening a participant with a diagnosis of gambling disorder according to the gold standard (full Pathological Gambling section of the DSM-5) as a problem gambler, using the screening tool tested.
Sp: Specificity; probability of screening a participant who does not have a diagnosis of gambling disorder according to the gold standard (full Pathological Gambling section of the DSM-5), as a NPG with the screening tool tested.
PPV: Positive Predictive Value; probability of a participant who is screened as a problem gambler using the tested screening tool, having a diagnosis of gambling disorder according to the gold standard (full Pathological Gambling section of the DSM-5).
NPV: Negative Predictive Value; probability of a participant who is screened as a NPG using the tested screening tool, not having a diagnosis of gambling disorder according to the gold standard (full Pathological Gambling section of the DSM-5).
ICR: Incorrect Classification Rate: proportion of misclassified gamblers (problem gamblers according to the gold standard who are classified as NPG using the tested screening tool and NPG according to the gold standard who are classified as problem gamblers using the tested screening tool).
YI, Se, and Sp are estimated values of the Youden Index, sensitivity and specificity obtained from the JEU sample excluding PGSTs.
PPV, NPV, and ICR were approximated using estimated values of Se and Sp in the JEU sample excluding PGSTs, and then re-computed to take into account the French prevalence rate of Pathological Gambling in the population of past-year gamblers (aPPV: adjusted PPV; aNPV: adjusted NPV; and aICR: adjusted ICR), that is 2.76% of at risk or excessive gambling (Costes et al., 2011).
Lifetime history of PG = subjective evaluation by the gambler whether he had or not a lifetime history of problem gambling.
[a]Three equivalents of interview-based Brief Screening Tools (Lie/Bet, BBGS, and NODS-CLiP).

The RSPG-I is composed of two criteria from the gambling disorder section of the DSM: the increasing bets criterion and the loss of control criterion. This version exhibited good screening properties and was especially efficient for screening gamblers with a diagnosis of gambling disorder (sensitivity of 0.95). Even if we selected the best combinations first based on the best sensitivity, the other properties remained at good levels. In particular,

Table 4. Screening properties of the best combinations of self-report items, and comparison with one equivalent of Brief Screening Tool (one-item SPG)

	Within the sample without PGSTs ($n = 425$)						Re-computation with national prevalence		
	YI	Se	Sp	PPV	NPV	ICR	aPPV	aNPV	aICR
Top 10 combinations of self-report items									
2-Item combinations									
SOGS 6 (past 12 months history of PG) and GABS 5 (losing track of time)	0.471	0.959	0.512	0.448	0.968	35.8%	0.053	0.998	47.6%
SOGS 4 (chasing) and GABS 5 (losing track of time)	0.447	0.959	0.488	0.436	0.966	37.4%	0.050	0.998	49.9%
SOGS 2 (max. bet in 1 day >€180) and SOGS 7 (gamble more than intended)	0.437	0.959	0.478	0.431	0.966	38.2%	0.050	0.998	50.9%
SOGS 8 (criticisms from others) and GABS 5 (losing track of time)	0.428	0.950	0.478	0.429	0.959	38.4%	0.049	0.997	50.9%
SOGS 10 (perceived inability to stop) and GABS 5 (losing track of time)	0.425	0.950	0.475	0.427	0.959	38.6%	0.049	0.997	51.2%
3-Item combinations									
SOGS 4 (chasing), SOGS 9 (guiltiness), and SOGS 10 (perceived inability to stop)	*0.588*	*0.950*	*0.637*	*0.519*	*0.969*	*27.1%*	*0.069*	*0.998*	*35.4%*
SOGS 2 (max. bet in 1 day >€180), SOGS 5 (wrongly claim to have won), and SOGS 11 (hiding gambling evidences)	0.578	0.950	0.627	0.512	0.968	27.8%	0.067	0.998	36.4%
SOGS 2 (max. bet in 1 day >€180), SOGS 9 (guiltiness), and SOGS 13 (arguments)	0.567	0.950	0.617	0.506	0.968	28.6%	0.066	0.998	37.4%
SOGS 2 (max. bet in 1 day >€180), SOGS 9 (guiltiness), and SOGS 16e (credit cards)	0.567	0.950	0.617	0.506	0.968	28.6%	0.066	0.998	37.4%
SOGS 2 (max. bet in 1 day >€180), SOGS 9 (guiltiness), and SOGS 15 (losing time from work)	0.564	0.950	0.614	0.504	0.968	28.8%	0.065	0.998	37.7%
Equivalents of brief screening tools[a]									
Item 6 from the SOGS (past 12 months history of PG) = equivalent of the one-item SPG (Rockloff et al., 2011)	0.603	0.752	0.851	0.675	0.893	17.8%	0.125	0.992	15.2%

Notes. Only combinations with the highest sensitivity (Se ≥ 0.95) were first selected, and the obtained list of combinations was then graded according to the highest Youden Index (YI). The retained combination is emphasized in italic font.
For all combinations, the screening of a gambling problem was based on at least one "yes" answer.
SOGS# and GABS#: SOGS and GABS item number.
YI: Youden Index = Sensitivity + Specificity − 1; performance of a diagnostic test, independently of the disorder's prevalence.
Se: Sensitivity; probability of screening a participant with a diagnosis of gambling disorder according to the gold standard (full Pathological Gambling section of the DSM-5) as a problem gambler, using the screening tool tested.
Sp: Specificity; probability of screening a participant who does not have a diagnosis of gambling disorder according to the gold standard (full Pathological Gambling section of the DSM-5), as a NPG with the screening tool tested.
PPV: Positive Predictive Value; probability of a participant who is screened as a problem gambler using the tested screening tool, having a diagnosis of gambling disorder according to the gold standard (full Pathological Gambling section of the DSM-5).
NPV: Negative Predictive Value; probability of a participant who is screened as a NPG using the tested screening tool, not having a diagnosis of gambling disorder according to the gold standard (full Pathological Gambling section of the DSM-5).
ICR: Incorrect Classification Rate: proportion of misclassified gamblers (problem gamblers according to the gold standard who are classified as NPG using the tested screening tool and NPG according to the gold standard who are classified as problem gamblers using the tested screening tool).
YI, Se, and Sp are estimated values of the Youden Index, sensitivity and specificity obtained from the JEU sample excluding PGSTs.
PPV, NPV, and ICR were approximated using estimated values of Se and Sp in the JEU sample excluding PGSTs and then re-computed to take into account the French prevalence rate of Pathological Gambling in the population of past-year gamblers (aPPV: adjusted PPV; aNPV: adjusted NPV; and aICR: adjusted ICR), that is 2.76% of at risk or excessive gambling (Costes et al., 2011).
[a]The equivalent of the self-report Brief Screening Tool (One-item SPG).

the RSPG-I showed good diagnosis efficiency (83.1%), in the same range as the NODS-CLiP (86.4% in the development study) (Volberg et al., 2011). The RSPG-I combines two crucial elements of addiction: tolerance and loss of control. For a long time, the definition of addiction emphasized tolerance and withdrawal as key components of drug dependence, which was only considered as a physical dependence (American Psychiatric Association, 1980). Since version IV of the DSM (American Psychiatric Association, 2000), symptoms of tolerance and withdrawal

Table 5. Classification accuracy of the Rapid Screener for Problem Gambling (RSPG) compared to gambling disorder DSM-5 diagnosis and severity, in the JEU cohort excluding PGSTs ($n = 425$)

			Current severity		
	GD−	GD+	Mild (4–5)	Moderate (6–7)	Severe (8–9)
Total *N*	301	124	57	44	23
% of sample	70.8%	29.2%	46.0%	35.5%	18.5%
N RSPG-I+	66	118	52	43	23
% RSPG-I+	21.9%	95.2%	91.2%	97.7%	100%
N RSPG-SA+	107 (MD = 6)	115 (MD = 3)	51 (MD = 1)	41 (MD = 2)	23 (MD = 0)
% RSPG-SA+	36.3%	95.0%	91.1%	97.6%	100%

Note. GD = diagnosis of a gambling disorder according to DSM-5; GD−/GD+ = without a gambling disorder/with a gambling disorder; RSPG−I+ = positive screening of a gambling problem according to the Rapid Screener for Problem Gambling – Interview; RSPG−SA+ = positive screening of a gambling problem according to the Rapid Screener for Problem Gambling – Self-Assessment; MD = missing data.

Table 6. Final Rapid Screener for Problem Gambling (RSPG) – English version

RSPG-I – Interview version (for a use in telephone-based or in face-to-face surveys)	
Have you had a gambling practice over the past 12 months?	
Yes	Continue the interview
No	Stop the interview – 0 point
During the past 12 months, have you needed to gamble with increasing amounts of money in order to achieve the desired excitement?	
Yes	1 point
No	0 point
During the past 12 months, have you made repeated unsuccessful efforts to control, cut back or stop gambling?	
Yes	1 point
No	0 point
Scoring and interpretation:	
Score of 0: don't need a deeper interview ⇒ no gambling disorder	
Score greater or equal to 1: conduct a deeper interview to establish a final diagnosis of a gambling disorder	
RSPG-SA – Self-assessment version (for a use in internet-based surveys)	
Have you had a gambling practice over the past 12 months?	
Yes	Continue the questionnaire
No	Stop the questionnaire – 0 point
When you gamble, how often do you go back another day to win back money you have lost?	
Never	0 point
Some of the time (less than half the time I lose)	0 point
Most of the time I lose	1 point
Every time I lose	1 point
Have you ever felt guilty about the way you gamble, or what happens when you gamble?	
Yes	1 point
No	0 point
Have you ever felt like you would like to stop betting money on gambling, but didn't think you could?	
Yes	1 point
No	0 point
Scoring and interpretation:	
Score of 0: don't need a deeper interview ⇒ no gambling disorder	
Score greater or equal to 1: conduct a deeper interview to establish a final diagnosis of a gambling disorder	

are no longer mandatory for a diagnosis of addiction, and behavioral symptoms (e.g., repeated but unsuccessful attempts to cut down or control the drug consumption or gambling practice, consuming drug or gambling more than intended, etc.) are now as important as the symptoms of physical dependence. The RSPG-I has thus the advantage of combining symptoms of both physical (tolerance) and behavioral (loss of control) dependence. Moreover, loss of control was one of the key elements of the definition of addictions proposed by Aviel Goodman (1990). This definition was especially well suited for behavioral addiction, and pathological gambling was mentioned as one of the disorders that corresponded very well to addiction criteria. Thus, the RSPG-I is fairly consistent with the reconciliation

Table 7. Final Rapid Screener for Problem Gambling (RSPG) – French version

RSPG-I – version Interview (pour les enquêtes téléphoniques ou en face-à-face)	
Avez-vous eu une pratique de jeu au cours des 12 derniers mois?	
Oui	Continuer l'interview
Non	Arrêter l'interview – 0 point
Durant les 12 derniers mois, avez-vous eu besoin de jouer avec des sommes d'argent de plus en plus élevées pour atteindre l'état d'excitation désiré?	
Oui	1 point
Non	0 point
Durant les 12 derniers mois, avez-vous fait des efforts répétés mais infructueux pour contrôler, réduire ou arrêter la pratique du jeu?	
Oui	1 point
Non	0 point
Cotation et interprétation:	
Score de 0: pas de besoin d'une interview plus poussée ⇒ pas de trouble lié au jeu	
Score supérieur ou égal à 1: faire une interview plus poussée pour établir un diagnostic final de trouble lié au jeu	
RSPG-SA – version Auto-Evaluation (pour les enquêtes par Internet)	
Avez-vous eu une pratique de jeu au cours des 12 derniers mois?	
Oui	Continuer le questionnaire
Non	Arrêter le questionnaire – 0 point
Lorsque vous avez joué au cours des 12 derniers mois, combien de fois êtes-vous retourné au jeu un autre jour pour vous refaire, c'est-à-dire pour regagner l'argent perdu auparavant?	
Jamais	0 point
Quelquefois (moins de la moitié des fois où j'ai perdu)	0 point
La plupart des fois où je perds	1 point
A chaque fois que je perds	1 point
Au cours des 12 derniers mois, vous êtes-vous déjà senti coupable à cause de la façon dont vous jouez ou à cause de ce qui se produit lorsque vous jouez?	
Oui	1 point
Non	0 point
Au cours des 12 derniers mois, avez-vous envisagé d'arrêter de jouer tout en pensant que vous en étiez incapable?	
Oui	1 point
Non	0 point
Cotation et interprétation:	
Score de 0: pas de besoin d'une interview plus poussée ⇒ pas de trouble lié au jeu	
Score supérieur ou égal à 1: faire une interview plus poussée pour établir un diagnostic final de trouble lié au jeu	

of gambling disorder and substance use disorders made in the DSM-5 (American Psychiatric Association, 2013). Another element of consistency with the DSM-5 is the absence in the RSPG-I of the illegal acts criterion, which is no longer considered for the diagnosis. Although this criterion was taken into account alongside other criteria to test all possible combinations, it has not been retained in the final version of the rapid screener. Toce-Gerstein and colleagues (Toce-Gerstein, Gerstein, & Volberg, 2003) identified the items 2 and 3 of the DSM-IV as being more present in pathological gamblers (with 5–7 positive criteria in DSM-IV). These items could thus indicate a certain level of severity of the gambling practice beyond the mere excessive use, which seems relevant with their use for a rapid screening of problem gambling.

The RSPG-SA is composed of three items from the SOGS (Lesieur & Blume, 1987): the chasing item, the guiltiness item, and the perceived inability to stop item. Despite of its self-report design, it displayed good screening properties. The RSPG-SA provided relevant screening of gamblers with a diagnosis of gambling disorder (sensitivity of 0.95) and showed correct diagnosis efficiency (72.9%). The chasing behavior had been demonstrated to be one of the DSM criteria, which discriminates best between community gamblers and gamblers in treatment (Stinchfield, Govoni, & Frisch, 2005). Moreover, the chasing item from the SOGS was found to be present for almost only probable pathological gamblers (SOGS score ≥ 5) in a telephone-based community survey (Cox, Kwong, Michaud, & Enns, 2000). Even if it was not the most frequently endorsed item, it seemed to be able to discriminate well between non-problem and problem gamblers, which seems relevant with its inclusion in the RSPG-SA. In the same study, guiltiness and inability to stop belonged to the most endorsed items in probable pathological gamblers. Compared to the chasing item (which rather refers to an externalized component of addiction), guiltiness and perceived inability to stop gambling could refer to more internalized components of gambling addiction. The presence of the item "guiltiness" in the self-report rapid screener is especially very interesting, because this notion is often forgotten from other problem gambling scales, despite of its clinical relevance. As argued by Flanagan (2013), the life of the addict is a source of both guilt and shame. Guiltiness does reflect the reactive attitude to failure to control gambling, as an intra-personal feeling of being unsuccessful both to control the behavior and to achieve the expected life made impossible because of the addiction. As this is not a necessary element to diagnose gambling disorder, but rather a psychological state of mind,

it is rarely investigated in problem gambling screening scales. But the inclusion of this aspect of gambling addiction in the RSPG is not surprising for the self-report version (who can be better suited than the gambler himself to assess his own state of mind?). It is also an indication of more severe problems, because guiltiness often appears when damages are already here and noticeable by others. Finally, perceived inability to stop gambling is rarely explored in gambling-related scales. Yet this dimension is a major clinical element characterizing problem gambling. For example, in a study exploring gambling-related cognitions, the perceived inability to stop gambling accounted for the highest variance in SOGS score and displayed a high discriminant function for identifying probable problem gamblers (Raylu & Oei, 2004). This dimension is indeed another indirect way to explore the loss of control over gambling, which could lead to low self-efficacy and the conviction that it is impossible to change the behavior. For a self-report measure, the recognition of a certain powerlessness to control the behavior may identify precociously problem gamblers long before important damage (which are often a focus of gambling problems related scales) settle down. The RSPG-SA has thus the advantage of combining both externalized and internalized components of addiction, and behavioral (chasing), emotional (guiltiness), and cognitive (perceived inability to stop) aspects.

A secondary objective of this study was to compare the RSPG to equivalents of the four major brief screening scales previously validated and widely used (one-item SPG, Lie/Bet, BBGS, and NODS-CLiP), especially in a sample with a good representation of both problem and non-problem gamblers. The equivalents were each better than the two versions of the RSPG in at least one screening property, depending on the preliminary choice made by the authors at the time of development of these scales. However, our strategy to select the best combinations based on both sensitivity and YI had allowed us to obtain two measures that are good in almost all screening properties, with no deficient property. Moreover, comparing to equivalents, the RSPG-I does not include item 7 from the DSM (lying), while this item is included in the three main equivalents (NODS-CLiP, BBGS, and Lie/Bet). This item is probably more prone to denial, which is a major bias problem in screening measures in addictology. Moreover, when the gambler has no close relative to whom he can lie, what is the relevance of this item?

Both the two versions of the RSPG have the advantage of combining complementary aspects of addiction and thus of being relatively complete despite of their short duration.

This work is limited by some weaknesses. First, the final RSPG questionnaires were not validated alone, in a representative sample of gamblers. Second, the list of items tested was obviously not exhaustive, and CPGI (Ferris & Wynne, 2001) items could especially not be included, whereas it is currently the mostly used questionnaire for screening gambling problems (the SOGS was most commonly used during startup of this study in 2009). Third, this work was done in French only, so that it could be useful to confirm the interest of the RSPG in other languages (the proposed English-version had not been tested itself). Fourth, the diagnosis of a gambling problem was made based on the 10 criteria from the DSM-IV, because the DSM-5 was not published at the time of recruitment and baseline assessment (2009–2011). DSM-5 changes include: reclassification, renaming, and changes in diagnostic criteria and lowering of threshold for a diagnosis (Reilly & Smith, 2013). All these changes could have led to underestimate the prevalence of addiction compared to the DSM-IV. However, a recent study of the impact of DSM-5-related changes on prevalence rates and classification accuracy concluded that the new criteria yielded equivalent or slightly better classification accuracy in all comparisons and across all samples (Petry, Blanco, Stinchfield, & Volberg, 2013). Fifth, the comparison with existing tools is not completely reliable because equivalents are not formulated strictly in the same way as the original tools so gamblers could have responded differently. Moreover, some items were passed in interview (DSM), while these tools are all used in self-report (NODS-CLiP, BBGS, Lie/Bet, and one-item, SPG). It is noteworthy that one of the strengths of our procedure was to differentiate the two modes of completion, while the existing screening tools are often used interchangeably in self-report (pen-and-paper or internet-based surveys) and interview (mostly telephone-based surveys) studies. Another great strength of this work was that all combinations of items from available gambling scales were tested in a systematic manner, with no *a priori*. Finally, to our knowledge, this is the first mini-screener for problem gambling based on DSM-5 categorization. Finally, and as expected, PPV were very low when being adjusted to national prevalence. However, the adjusted VPN remains very good even with low national prevalence, as this is desired for a screening tool.

To conclude, we recommend to use the RSPG-SA/I for screening problem gambling in epidemiological surveys, with the version adapted for each purpose (RSPG-I for telephone-based or face-to-face survey, and RSPG-SA for pen-and-paper or internet-based surveys). This first triage of potential problem gamblers using the RSPG-SA/I must be supplemented by further assessment (Problem Gambling Research and Treatment Centre (PGRTC), 2011; Toce-Gerstein et al., 2009), as it may overestimate the proportion of problem gamblers (specificity = 0.78 or 0.64). However, a first triage has the great advantage of saving time and energy in large-scale screening for problem gambling.

Funding sources: This study was supported by both the joint support of the French Inter-departmental Mission for the fight against drugs and drug addiction (MILDT) and the French National Institute of Health and Medical Research (INSERM), as part of the call for research projects launched by these two organizations in 2007, and a grant from the French Ministry of Health (PHRC 2009 – RCB 2008-A01188-47). There were no constraints on publishing.

Authors' contribution: MGB and GCB designed the study, conducted literature search, wrote the protocol, and are responsible for the project management. JBH provided methodological advice and designed the statistical analysis

plan. BP and JBH conducted the statistical analysis. All other authors (including those mentioned in the JEU Group) contributed to include the participants in the study, to the development of the protocol, and to obtain funding sources. GCB wrote the first draft of the manuscript, and all authors gave feedback and approved the final manuscript.

Conflict of interest: MGB and GCB declare that the University Hospital of Nantes has received funding from the gambling industry (FDJ and PMU) in the form of a sponsorship, which supports the gambling section of the BALANCED Unit (the Reference Centre for Excessive Gambling). Scientific independence toward gambling industry operators is warranted. There were no constraints on publishing.

LR declares that the University of Paris Ouest Nanterre La Défense has received funding directly from gambling industry (FDJ and PMU) as part of other research contracts – this funding has never had any influence on this work. BP, JBH, MV, DM, MF, ICB, and AL declare that they have no conflicts of interest.

Acknowledgments: We wish to sincerely thank all the staff, who contributed to this study (JEU Group), for their valuable assistance and significant investment. Special thanks go to those who collected the data. This research was conducted at the initiative of and coordinated by the Clinical Investigation Unit *BALANCED* "BehaviorAL AddictioNs and ComplEx mood Disorders" of the University Hospital of Nantes, the sponsor of this study.

REFERENCES

American Psychiatric Association. (1980). *DSM-III. Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2000). *DSM-IV-TR. Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: American Psychiatric Association.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental health disorders: DSM-5* (5th ed.). Washington, DC: American Psychiatric Association.

Bouju, G., Hardouin, J. B., Boutin, C., Gorwood, P., Le Bourvellec, J. D., & Feuillet, F. (2014). A shorter and multidimensional version of the Gambling Attitudes and Beliefs Survey (GABS-23). *Journal of Gambling Studies, 30,* 349–367. doi:10.1007/s10899-012-9356-3

Breen, R. B., & Zuckerman, M. (1999). 'Chasing' in gambling behavior: Personality and cognitive determinants. *Personality and Individual Differences, 27,* 1097–1111. doi:10.1016/S0191-8869(99)00052-5

Challet-Bouju, G., Hardouin, J. B., Vénisse, J. L., Romo, L., Valleur, M., & Magalon, D. (2014). Study protocol: The JEU cohort study – Transversal multiaxial evaluation and 5-year follow-up of a cohort of French gamblers. *BMC Psychiatry, 14,* 226. doi:10.1186/s12888-014-0226-7

Costes, J. M., Pousset, M., Eroukmanoff, V., Le Nezet, O., Richard, J. B., & Guignard, R. (2011). Les niveaux et pratiques des jeux de hasard et d'argent en 2010. *Tendances, 77*, 1–8.

Cox, B. J., Kwong, J., Michaud, V., & Enns, M. W. (2000). Problem and probable pathological gambling: Considerations from a community survey. *Canadian Journal of Psychiatry, 45,* 548–553. doi:10.1177/070674370004500606

Ferris, J., & Wynne, H. (2001). *L'indice Canadien du jeu excessif: Rapport final [The Canadian problem gambling index: Final report]* (pp. 1–72). Ottawa, ON: Centre canadien de lutte contre l'alcoolisme et les toxicomanies.

Flanagan, O. (2013). The shame of addiction. *Frontiers in Psychiatry, 4,* 1–11. doi:10.3389/fpsyt.2013.00120

Gebauer, L., LaBrie, R., & Shaffer, H. J. (2010). Optimizing DSM-IV-TR classification accuracy: A brief biosocial screen for detecting current gambling disorders among gamblers in the general household population. *Canadian Journal of Psychiatry, 55,* 82–90. doi:10.1177/070674371005500204

Gerstein, D., Hoffman, J., Larison, C., Engelman, L., Murphy, S., & Plamer, A. (1999). Gambling Impact and Behavior Study. *Report of the National Gambling Impact Study Commission (NORC)*. Retrieved from http://www.norc.org/pdfs/publications/gibsfinalreportapril1999.pdf

Goodman, A. (1990). Addiction: Definition and implications. *British Journal of Addiction, 85,* 1403–1408. doi:10.1111/add.1990.85.issue-11

Grant, J. E., Steinberg, M. A., Kim, S. W., Rounsaville, B. J., & Potenza, M. N. (2004). Preliminary validity and reliability testing of a structured clinical interview for pathological gambling. *Psychiatry Research, 128,* 79–88. doi:10.1016/j.psychres.2004.05.006

Hodgins, D. C. (2004). Using the NORC DSM screen for gambling problems as an outcome measure for pathological gambling: Psychometric evaluation. *Addictive Behaviors, 29,* 1685–1690. doi:10.1016/j.addbeh.2004.03.017

Johnson, E. E., Hamer, R., Nora, R. M., Tan, B., Eisenstein, N., & Engelhart, C. (1997). The Lie/Bet Questionnaire for screening pathological gamblers. *Psychological Reports, 80,* 83–88. doi:10.2466/pr0.1997.80.1.83

Lejoyeux, M. (1999). Echelles de dépistage du jeu pathologique [Screening scales for pathological gambling]. *Neuro-psy, 14,* 67–71.

Lesieur, H. R., & Blume, S. B. (1987). The South Oaks Gambling Screen (SOGS): A new instrument for the identification of pathological gamblers. *The American Journal of Psychiatry, 144,* 1184–1188. doi:10.1176/ajp.144.9.1184

Petry, N. M., Blanco, C., Stinchfield, R., & Volberg, R. (2013). An empirical evaluation of proposed changes for gambling diagnosis in the DSM-5. *Addiction, 108,* 575–581. doi:10.1111/add.2013.108.issue-3

Problem Gambling Research and Treatment Centre (PGRTC). (2011). *Guideline for screening, assessment and treatment in problem gambling* (pp. 1–126). Clayton, Australia: Monash University.

Raylu, N., & Oei, T. P. (2004). The Gambling Related Cognitions Scale (GRCS): Development, confirmatory factor validation and psychometric properties. *Addiction, 99,* 757–769. doi:10.1111/add.2004.99.issue-6

Reilly, C., & Smith, N. (2013). The evolving definition of pathological gambling in the DSM-5. White paper of the National Center for Responsible Gaming, 1–6.

Rockloff, M. J., Ehrich, J., Themessl-Huber, M., & Evans, L. G. (2011). Validation of a one item screen for problem gambling. *Journal of Gambling Studies, 27,* 701–707. doi:10.1007/s10899-010-9232-y

Stinchfield, R., Govoni, R., & Frisch, G. R. (2005). DSM-IV diagnostic criteria for pathological gambling: Reliability, validity, and classification accuracy. *The American Journal of Psychiatry, 14,* 73–82. doi:10.1080/10550490590899871

Stinchfield, R., McCready, J., & Turner, N. (2012). A comprehensive review of problem gambling screens and scales for online self-assessment. Report prepared for the Ontario Problem Gambling Research Centre. Retrieved from http://greo.ca/sites/default/files/documents/Stinchfield%20et%20al%282012%29A_comprehensive_review_of_PG_self_assessment_tools.pdf

Thomas, S. A., Piterman, L., & Jackson, A. C. (2008). Problem gambling: What do general practitioners need to know and do about it? *Medical Journal of Australia, 189,* 135–136.

Toce-Gerstein, M., Gerstein, D. R., & Volberg, R. A. (2003). A hierarchy of gambling disorders in the community. *Addiction, 98,* 1661–1672. doi:10.1111/add.2003.98.issue-12

Toce-Gerstein, M., Gerstein, D. R., & Volberg, R. A. (2009). The NODS-CLiP: A rapid screen for adult pathological and problem gambling. *Journal of Gambling Studies, 25,* 541–555. doi:10.1007/s10899-009-9135-y

Tolchard, B., & Battersby, M. W. (2010). The Victorian Gambling Screen: Reliability and validation in a clinical population. *Journal of Gambling Studies, 26,* 623–638. doi:10.1007/s10899-009-9172-6

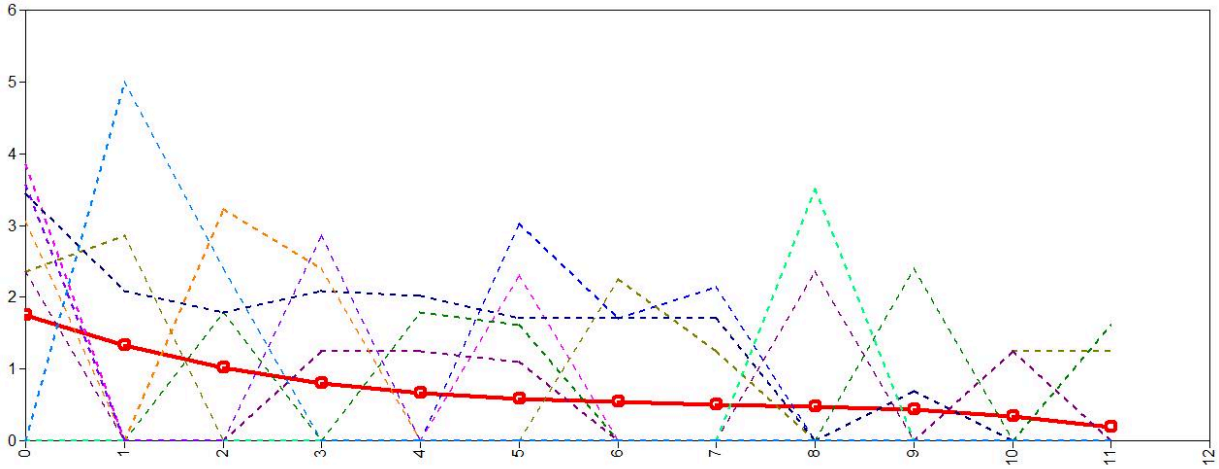
Volberg, R. A., Munck, I. M., & Petry, N. M. (2011). A quick and simple screening method for pathological and problem gamblers in addiction programs and practices. *The American Journal on Addictions/American Academy of Psychiatrists in Alcoholism and Addictions, 20,* 220–227. doi:10.1111/ajad.2011.20.issue-3

Winters, K. C., Specker, S., & Stinchfield, R. (2002). Measuring pathological gambling with The Diagnostic Interview for Gambling Severity (DIGS). In J. J. Marotta, J. A. Cornelius, & W. R. Eadington (Eds.), *The downside: Problem and pathological gambling* (pp. 143–148). Reno, NV: University of Nevada.

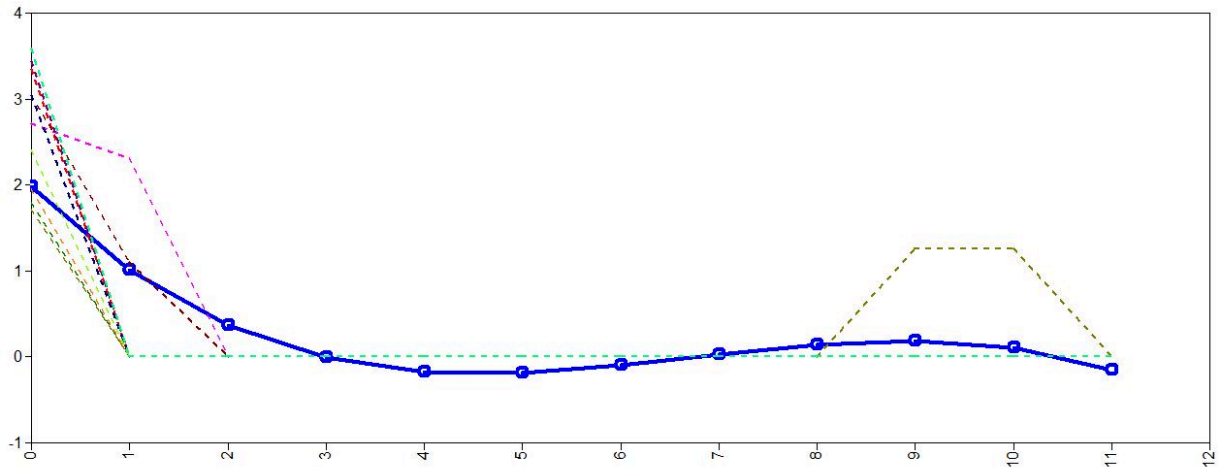
Youden, W. J. (1950). Index for rating diagnostic tests. *Cancer, 3,* 32–35. doi:10.1002/(ISSN)1097-0142

Annexe C

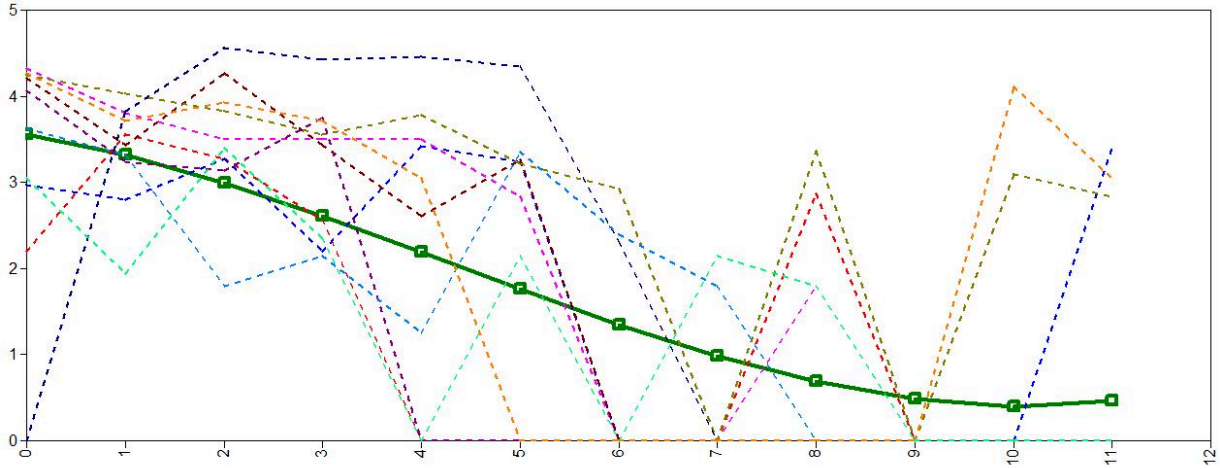
Représentation des trajectoires estimées moyennes et des trajectoires individuelles observées



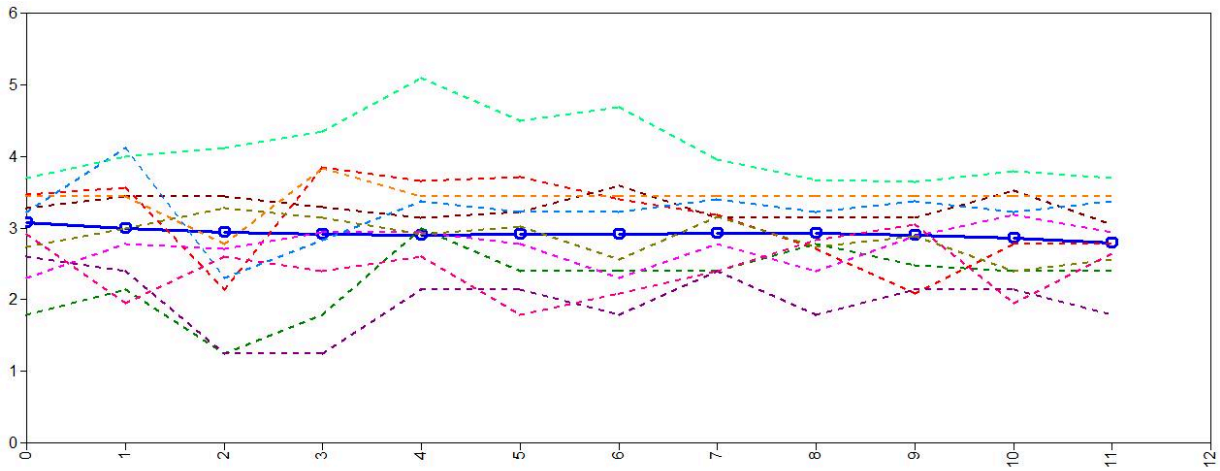
(a) Classe 1



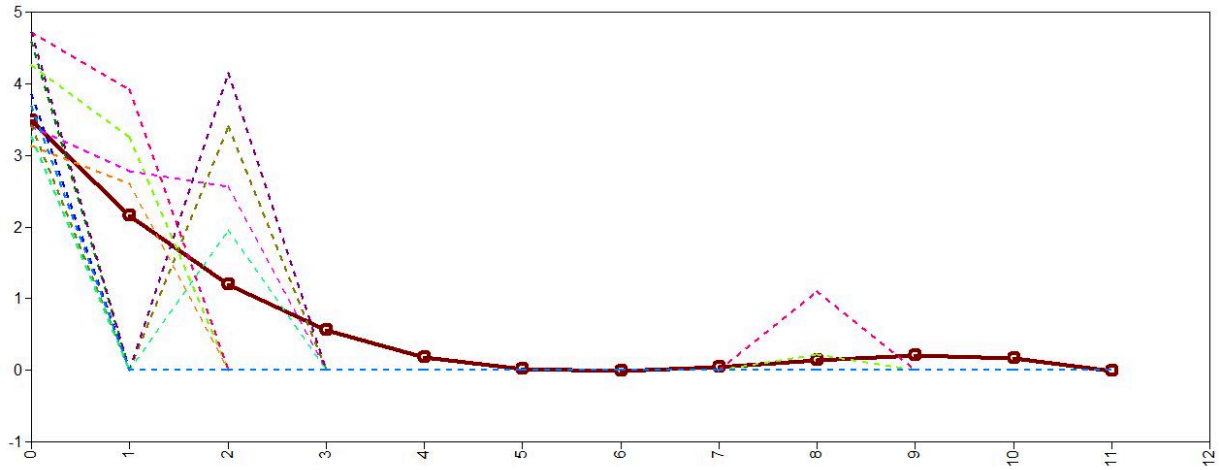
(b) Classe 2



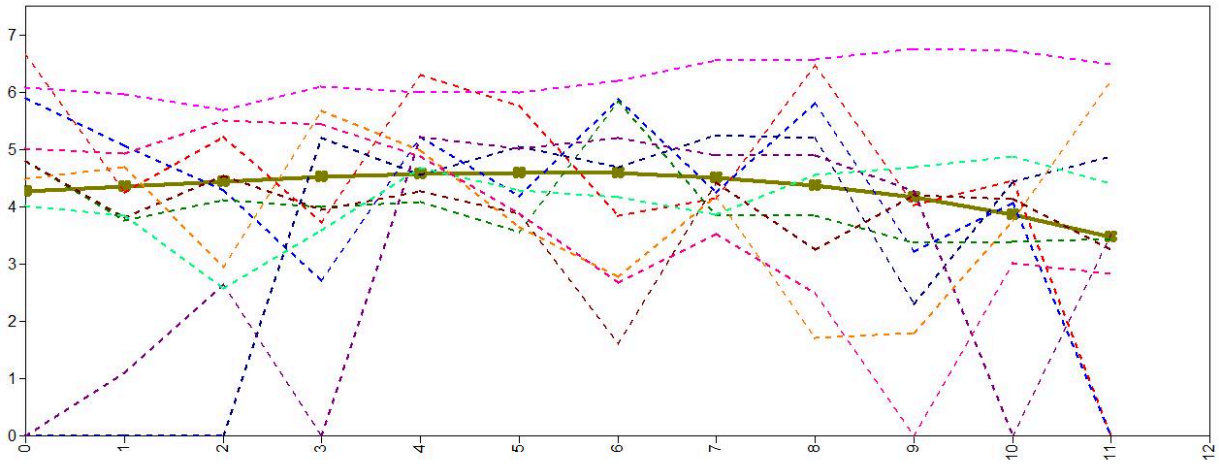
(c) Classe 3



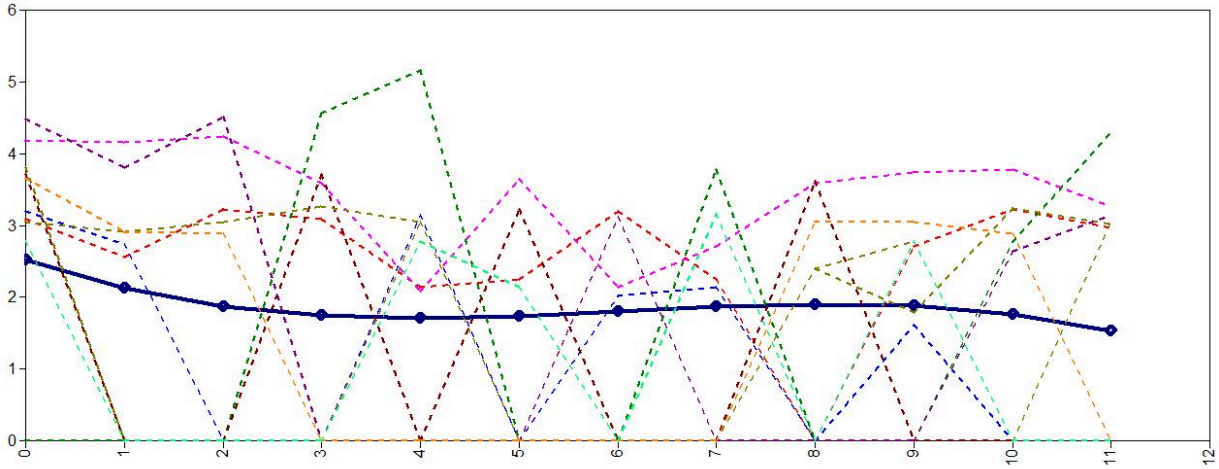
(d) Classe 4



(e) Classe 5



(f) Classe 6



(g) Classe 7

FIGURE C.-1 – Trajectoires estimées moyennes et trajectoires individuelles observées

UNIVERSITE BIOLOGIE BRETAGNE SANTE LOIRE

UNIVERSITÉ DE NANTES

Titre : Repérage des problèmes liés à la pratique des jeux de hasard et d'argent sur Internet

Mots clés : Jeux de hasard et d'argent ; Jeu problématique ; Prévention ; Jeux de hasard et d'argent sur Internet ; Comportement de jeu ; Modèle à classes latentes.

Résumé : La pratique des jeux de hasard et d'argent (JHA) est une activité de loisir, qui peut s'avérer problématique. En effet, certains joueurs peuvent perdre le contrôle sur leur pratique de jeu et subir des conséquences négatives d'ordre social, personnel ou professionnel. De plus, les personnes jouant à des JHA sur Internet seraient particulièrement à risque de développer des problèmes de jeu. Le repérage des pratiques de jeu problématiques, en particulier sur Internet, constitue une étape importante dans la mise en place de mesures de prévention. L'objectif principal de cette thèse était de réaliser une typologie des comportements de jeu en ligne pour pouvoir identifier des comportements de jeu à risque. Pour ce faire, une analyse en classes latentes multi-niveaux a été effectuée à partir des données d'activité de jeu de personnes possédant un compte sur le site d'un opérateur de jeu. Le modèle a permis de mettre en évidence des clusters de comportements de jeu mensuels distincts, et notamment des comportements de jeu pouvant être considérés comme problématiques. Nous avons également identifié des profils de joueurs dont certains pourraient correspondre à des joueurs à risque. En couplant ces résultats à des données rapportées par les joueurs grâce à des questionnaires, il serait possible de mettre en place des mesures de prévention ciblées selon les profils des joueurs.

Title : Screening of online gambling problems ....

Keywords : Gambling; Problem gambling; Prevention; Online gambling; Gambling behavior; Latent class models.

Abstract : Gambling is a leisure activity which can turn out to be problematic. Indeed, some gamblers can lose control over their gambling practice and suffer negative social, personal or professional consequences. Moreover, people who gamble online seem to be particularly at risk of gambling problems. Screening of problematic gambling practices, in particular on the Internet, is an important step in the setting of prevention measures. The main objective of the present thesis was to define a typology of online gambling behaviors, in order to identify risky gambling behaviors. A multilevel latent class cluster analysis was performed based on player account-based data of gamblers owning an account on a gambling operator website. The model highlighted distinct clusters of monthly gambling behaviors, including ones corresponding to likely problematic gambling behaviors. We also identified gamblers profiles, some of which could correspond to at risk gamblers. By linking these results with gambler reported data collected through questionnaires, it would be possible to set up targeted prevention measures, adapted to gamblers profile.