

Colloque Jeunes chercheurs CIQSS - Février 2007

*Imputer ou ne pas imputer des valeurs probables aux données manquantes.
Le cas de l'indice de sévérité du jeu pathologique*

Aline Drapeau

Richard Boyer

Géraldine Lo Siou.

Pourquoi se préoccuper des valeurs manquantes?

- Diminution de l'échantillon
 - Diminution de la puissance statistique
 - Diminution de la précision
- Biais de sélection

Dans l'ESCC-1.2,
(cycle 1.2 Enquête sur la santé dans les collectivités canadiennes)

24% des répondants ont une valeur manquante
pour l'indice de gravité des problèmes de jeu ...

Ça fait mal!

Pourquoi?

- Ils ont répondu spontanément qu'ils n'étaient pas des parieurs à la question
Au cours des 12 derniers mois, combien de fois avez-vous parié ou dépensé plus d'argent au jeu que vous en aviez l'intention?
- Connotation négative du mot 'parieur'?
Risque de biais de sélection

Indice canadien de jeu excessif (ICJE)

Définition du jeu excessif:

*Le jeu excessif est un comportement lié au jeu qui entraîne des conséquences négatives tant pour le joueur lui-même que pour les personnes de son réseau social ou pour la collectivité
(Ferris et Wynne, 2001)*

**Indice canadien de jeu excessif (ICJE)
(Ferris/Wynne, 2001)
Adaptation de l'ICJE pour l'ESCC-1.2**

- (1) Activités de jeu
 - Fréquence 13 items
 - Coût 1 item

- (2) Gravité des problèmes de jeu
 - 9 items

- (3) Corrélats
 - Antécédents familiaux 1 item
 - Association alcool/drogue 1 item

Mesure des activités de jeu dans l'ESCC-1.2

- 13 items de fréquence

- : loterie quotidienne/instant
- : autre loterie (6/49; fonds)
- : bingo
- : jeux de cartes (argent)
- : loterie vidéo hors casino
- : loterie vidéo au casino
- : poker, roulette au casino
- : Internet ou arcade
- : course de chevaux
- : loterie sportive
- : Bourse, actions
- : jeux d'habiletés
- : autres

- 8 choix de réponse

- : Tous les jours
- : 2-6 fois/semaine
- : 1 fois/semaine
- : 2-3 fois/mois
- : 1 fois/mois
- : 6-11 fois/année
- : 1-5 fois/année
- : Jamais

- 1 item de coût

(6 catégories: 1 à 1000\$)

- Période de référence : 1 an

Évaluation de la gravité des problèmes de jeu: 9 items

Comportements problématiques

- : Dépenser + pour même excitation
- : Rejouer pour se refaire
- : Emprunter ou vendre des biens
- : Penser qu'on a un problème de jeu
- : Parier plus qu'on peut se permettre

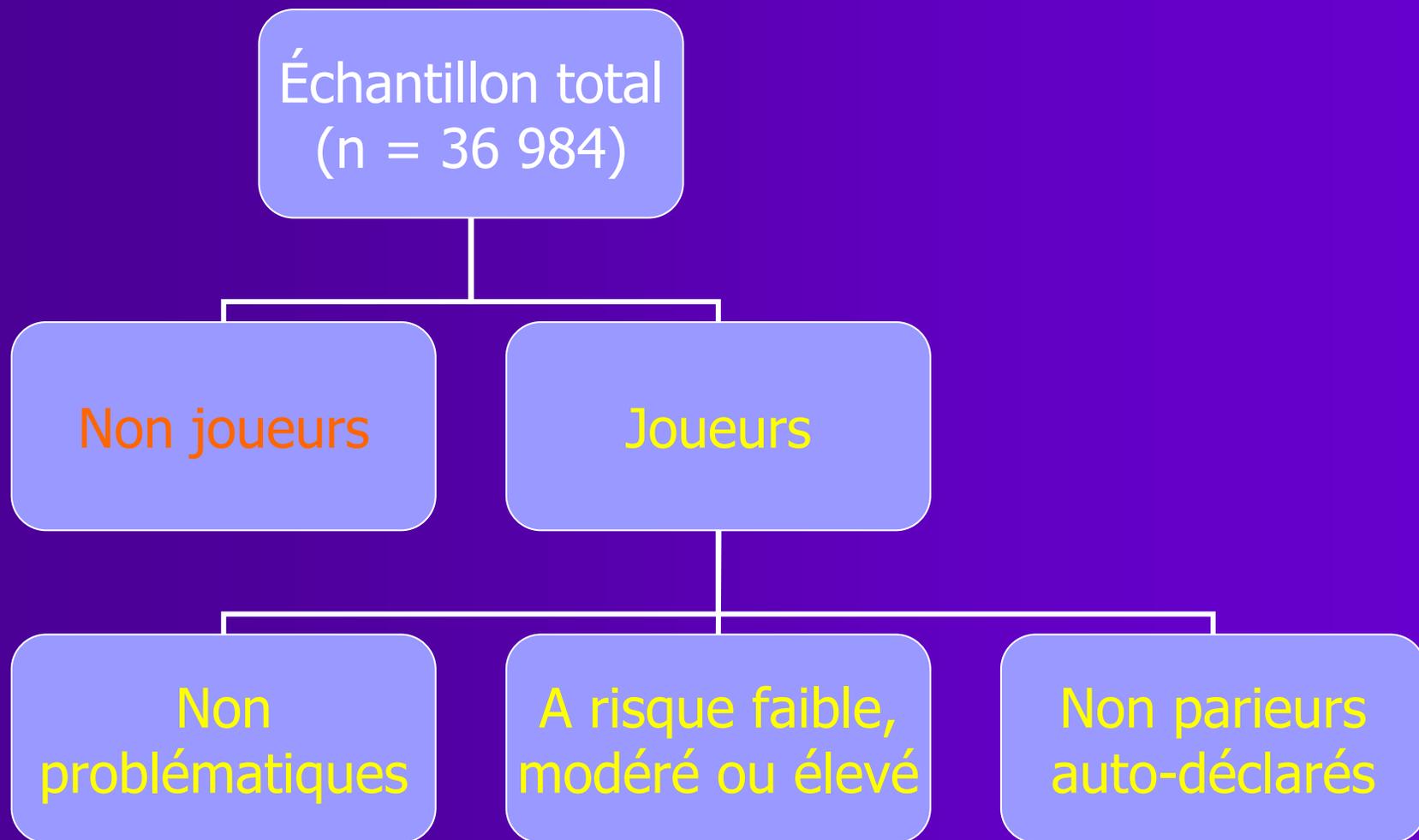
- 4 choix de réponse
 - : Jamais
 - : Quelques fois
 - : Plupart du temps
 - : Presque toujours

Conséquences négatives

- : Problèmes de santé
- : Être critiqué
- : Problèmes financiers
- : Sentiment de culpabilité

- Score potentiel 0-27
- Période de référence : 1 an
- Question-filtre:
Parier plus que prévu

Six catégories de joueur basées sur l'indice de gravité des problèmes de jeu (IGPJ)

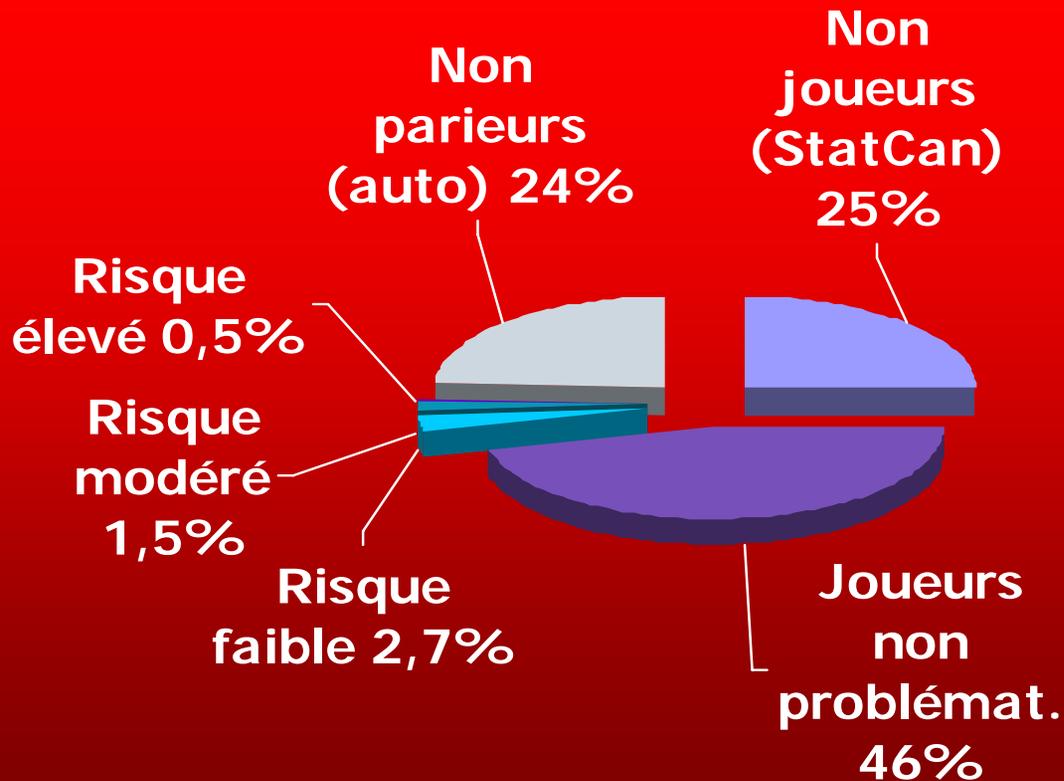


Six catégories de joueur basées sur l'indice de gravité des problèmes de jeu (IGPJ)

- **Non joueurs (n = 8930)**
Aucune activité de jeu de hasard dans les 12 derniers mois
- **Joueurs non problématiques selon StatCan (n = 17225)**
Répondu à l'IGPJ et obtenu un score de 0 à IGJP
Pas répondu à IGPJ (Parier 5 fois ou moins/année/activité)
- **Risque faible (Score de 1 ou 2 à IGPJ) (n = 1032)**
- **Risque modéré (Score de 3 à 7 à IGPJ) (n = 551)**
- **Risque élevé (Score de 8 et plus à IGPJ) (n = 174)**
- **Non parieurs auto-déclarés (n = 8962)**
Parié plus de 5 fois/année/activité
Répondu *je ne suis pas parieur* à la question-filtre

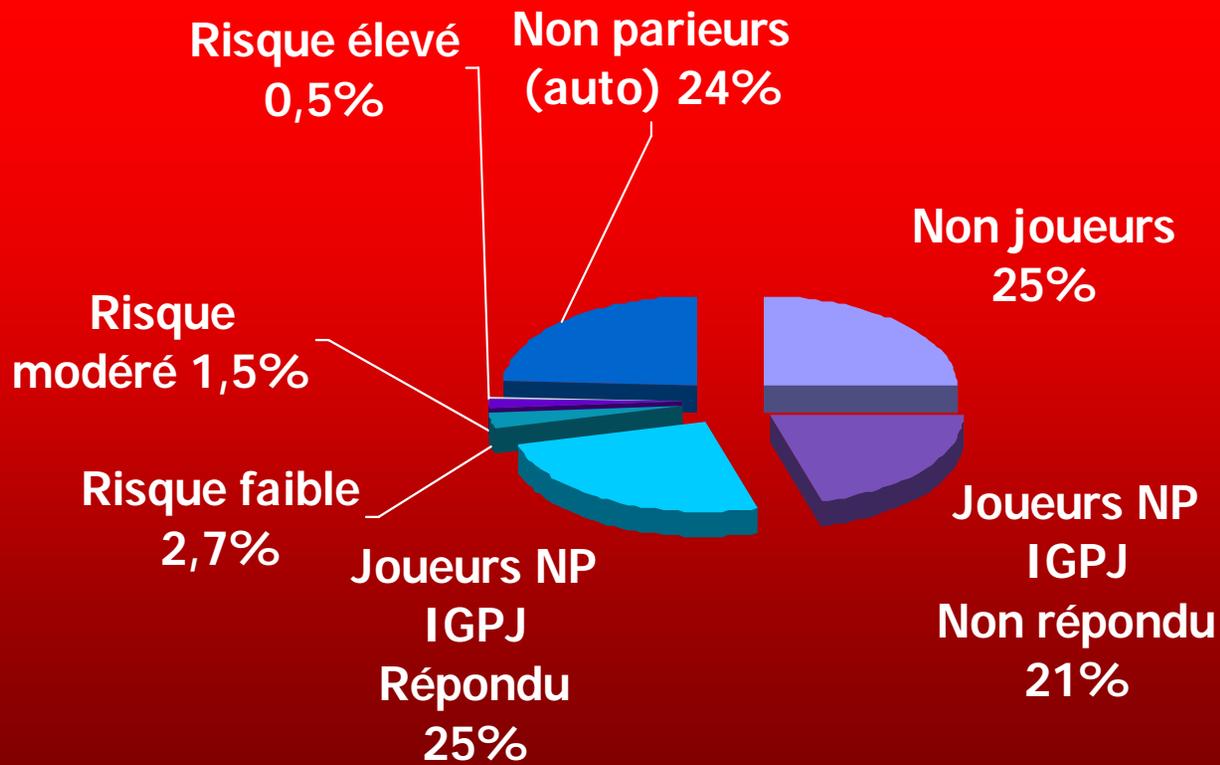
Indice de gravité des problèmes de jeu

Distribution initiale



Indice de gravité des problèmes de jeu

Distribution initiale décortiquée



Que faire avec les 24% de répondants qui ont déclaré ne pas être des parieurs et qui ont une valeur manquante à l'IGPJ?

■ Scénario # 1 Aucune imputation

Respect des limites des données
Perte de puissance statistique
Biais de sélection

■ Scénario # 2 Imputation StatCan

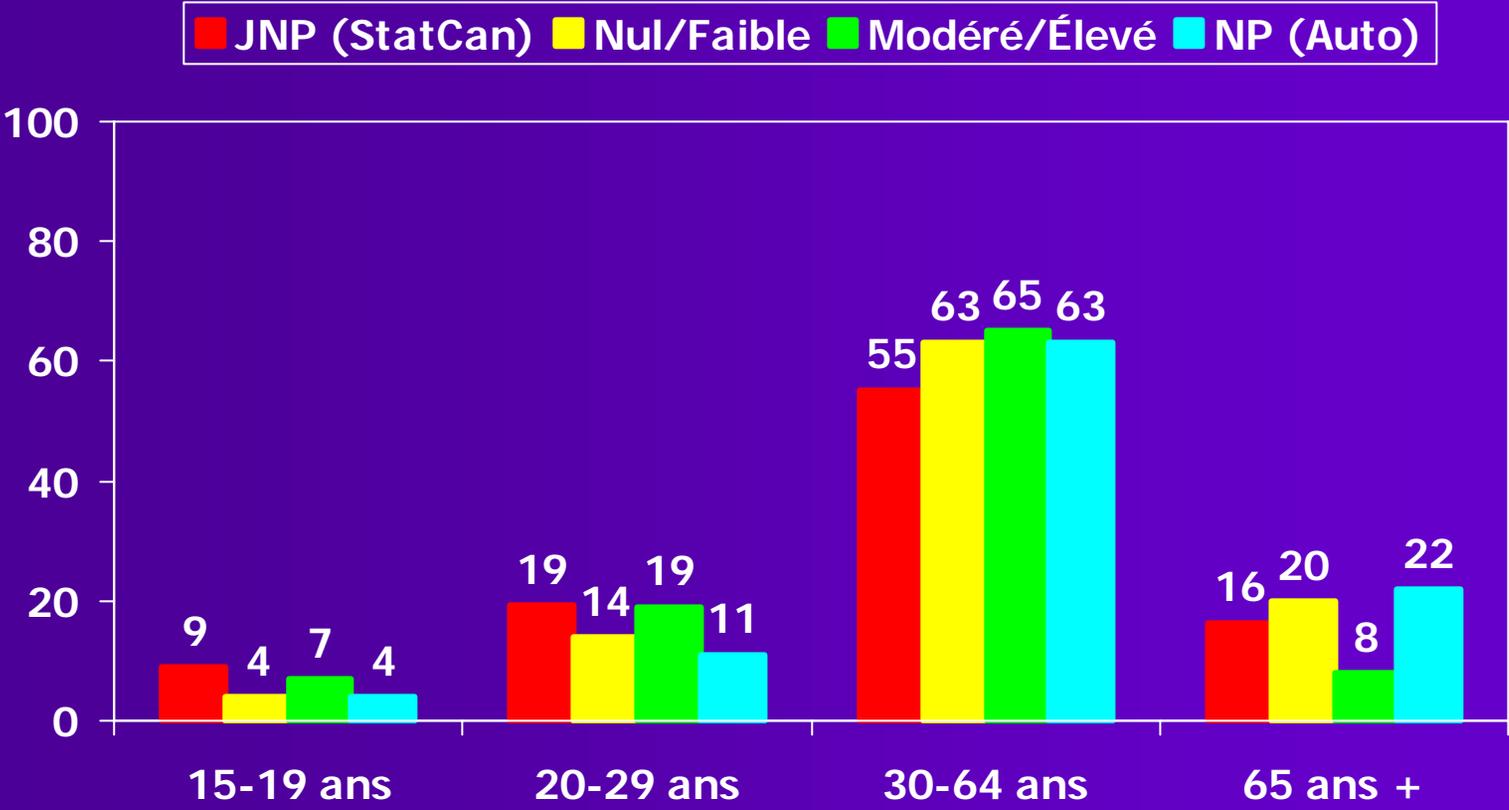
Non parieurs auto-déclarés = classés JNP
Artéfact méthodologique?
Comparabilité des données

■ Scénario # 3 Imputation de valeurs probables

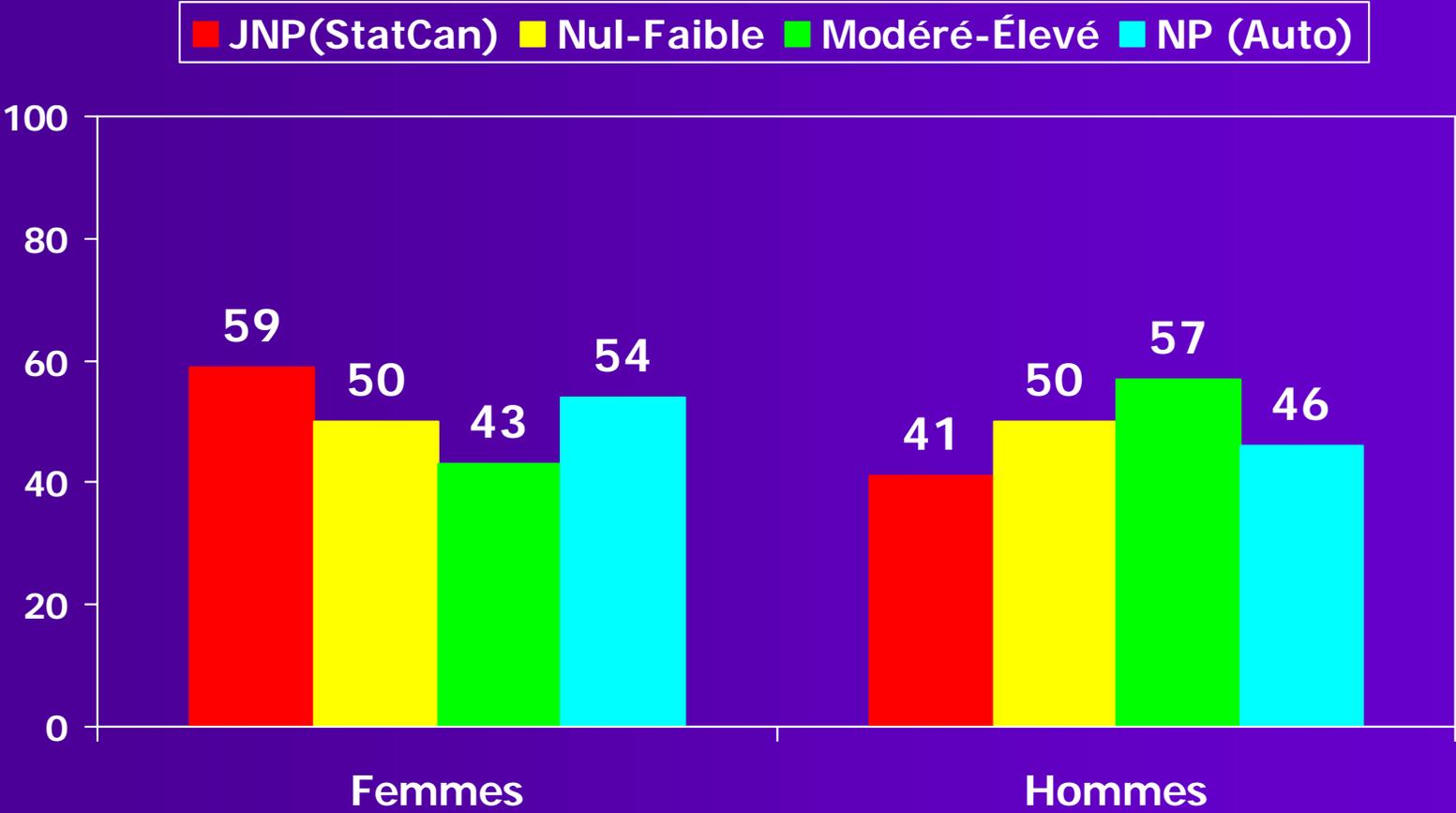
Artéfact méthodologique?
Plus valide que l'imputation StatCan?
Perte de comparabilité

Qui sont ces non-parieurs auto-déclarés?

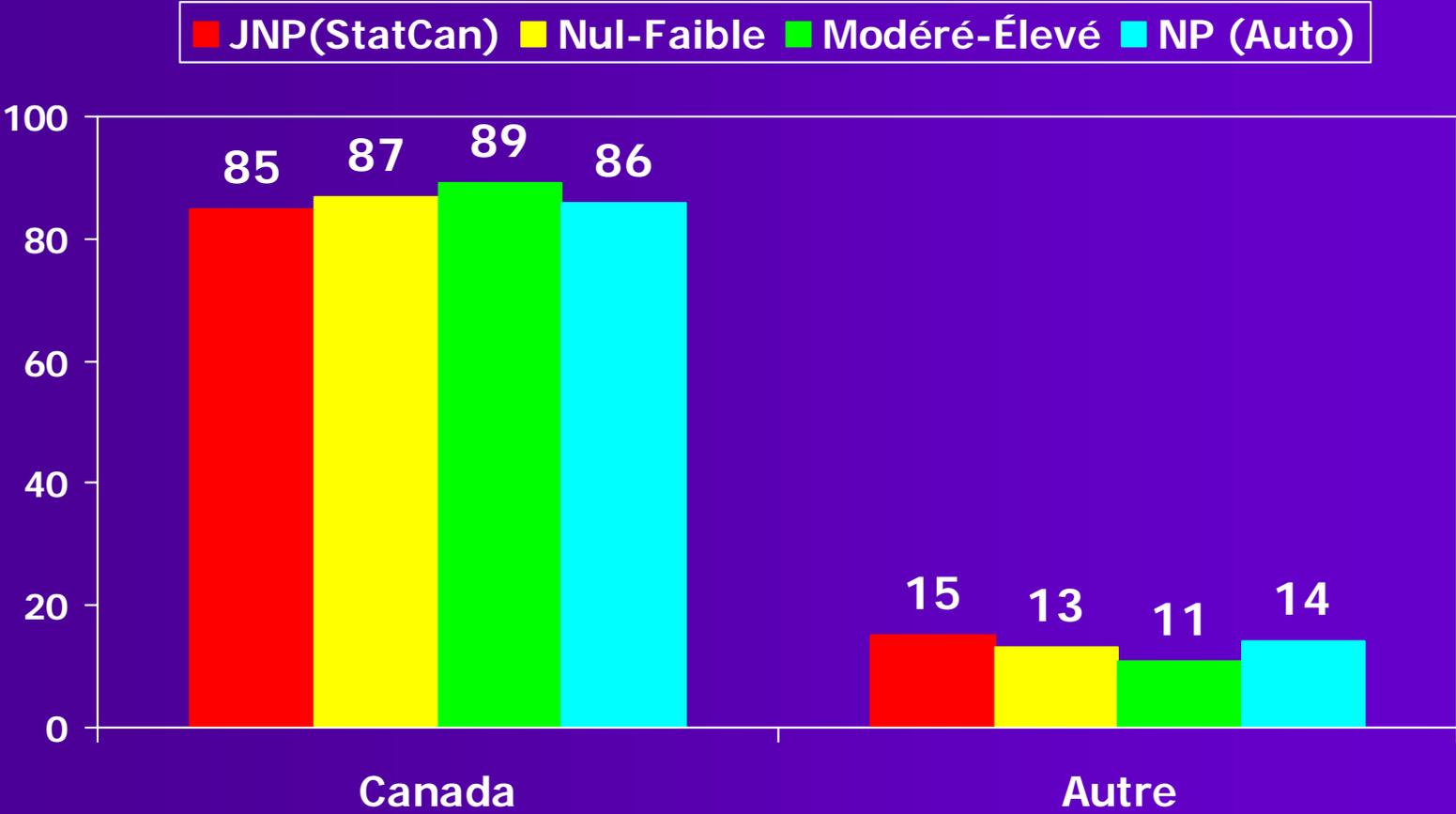
Catégories de joueurs selon le groupe d'âge (%)



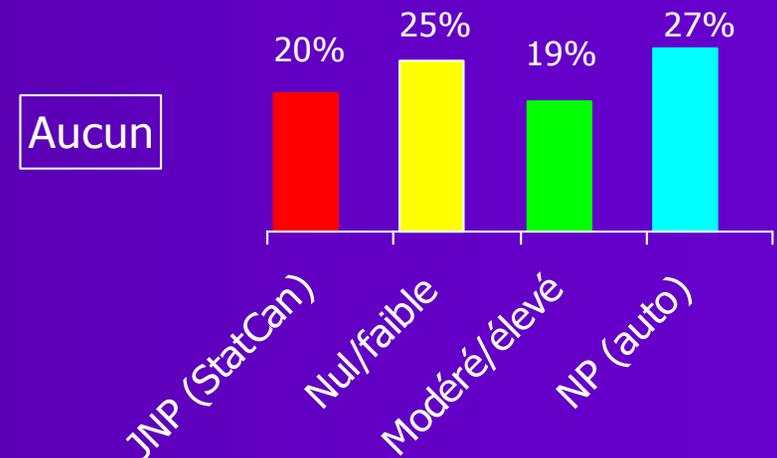
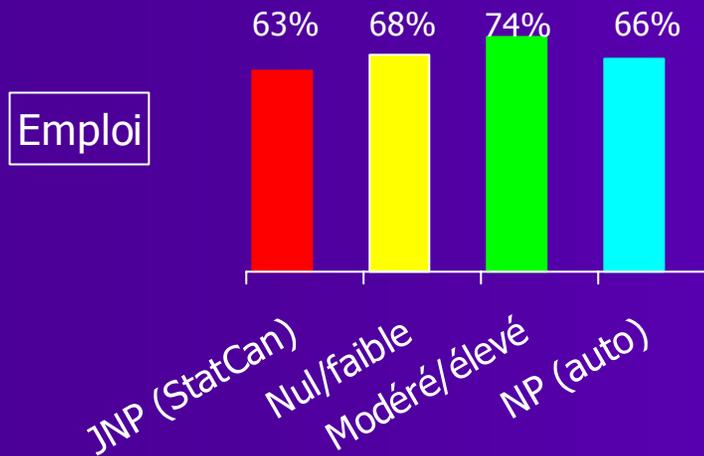
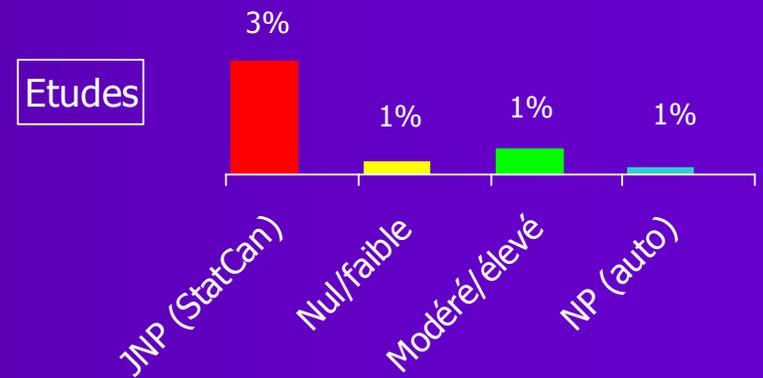
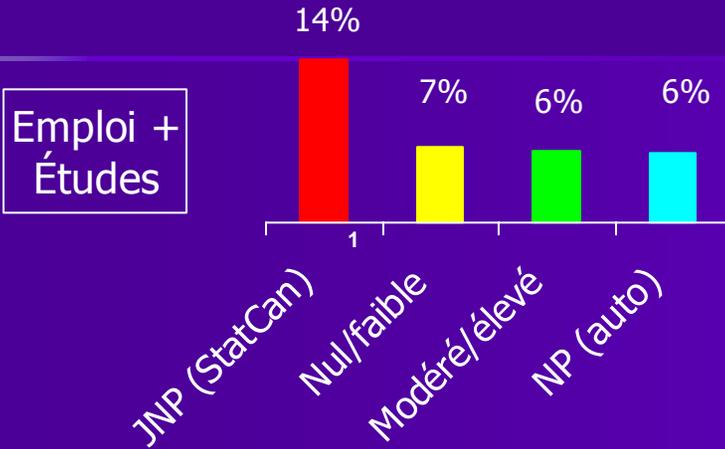
Catégories de joueurs selon le sexe (%)



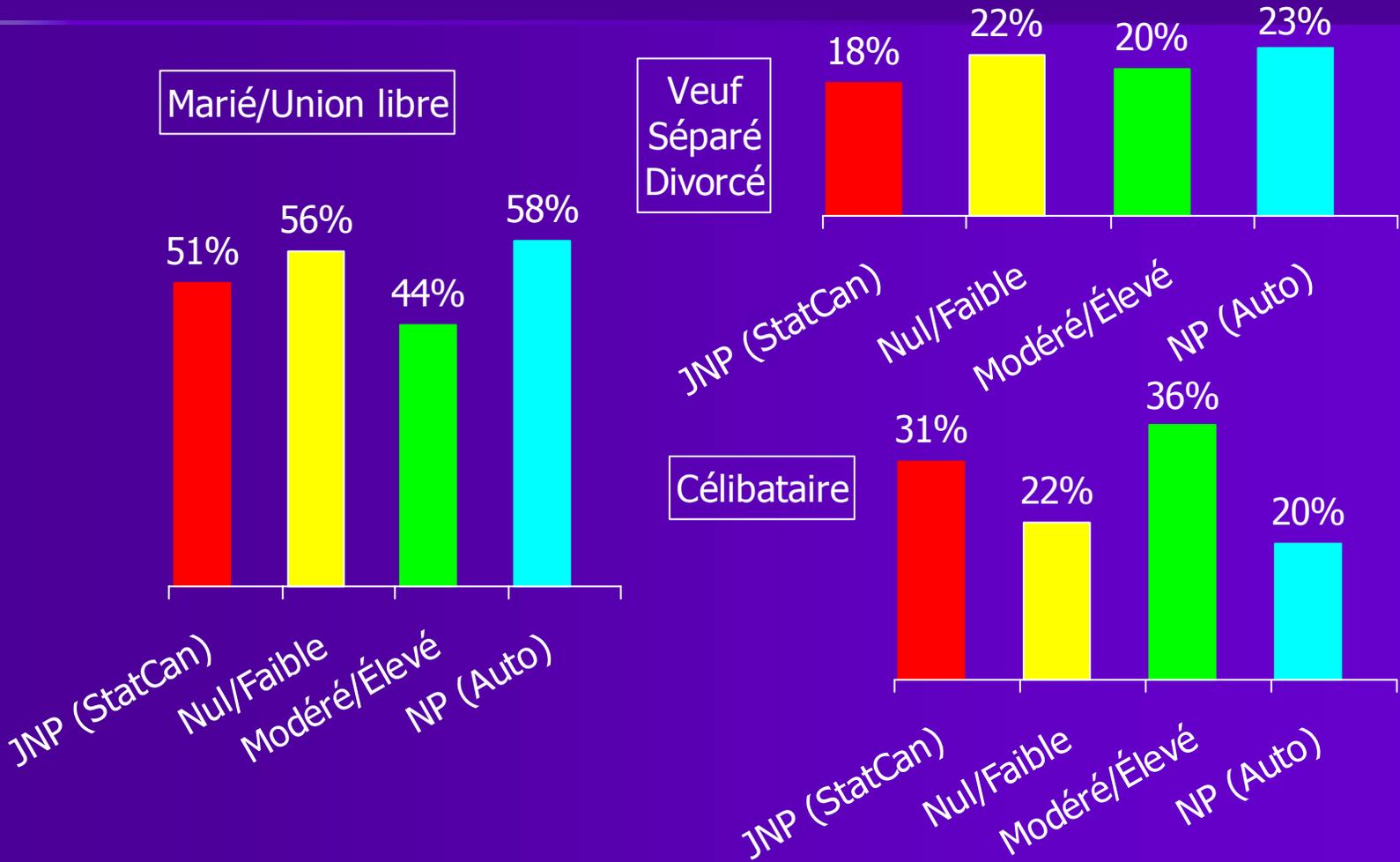
Catégories de joueurs selon le pays de naissance (%)



Catégories de joueurs selon l'occupation (%)

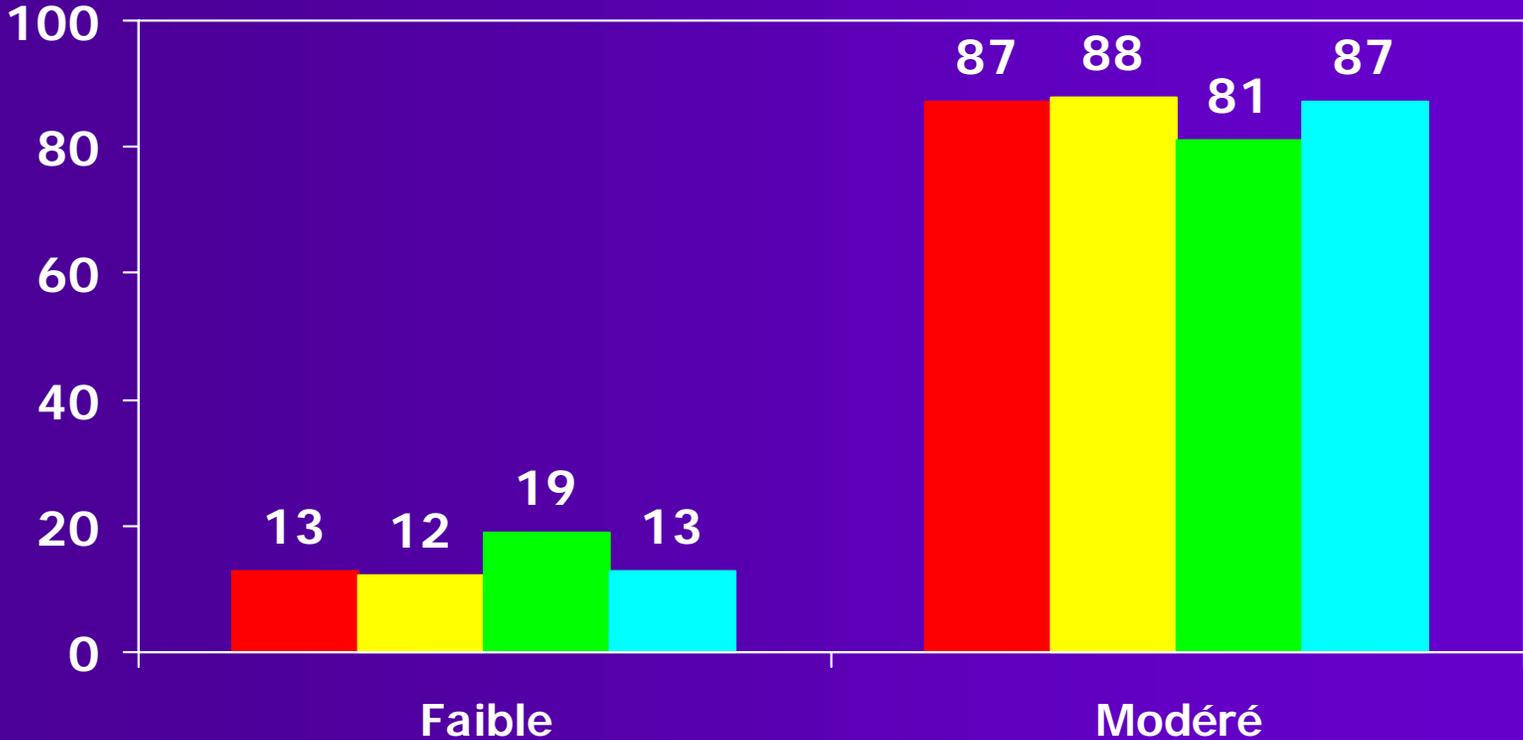


Catégories de joueurs selon le statut marital (%)

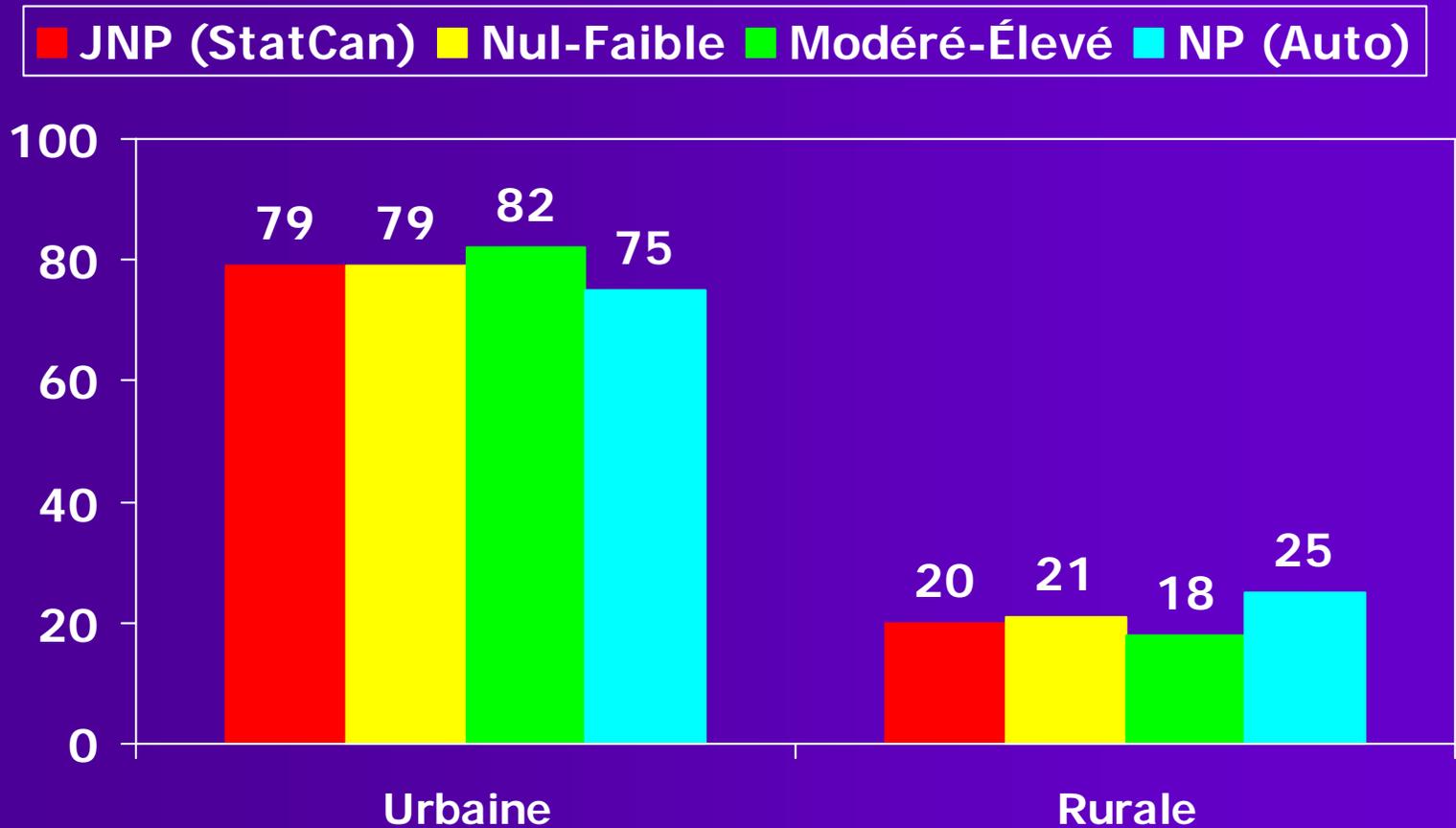


Catégories de joueurs selon le revenu du ménage (%)

■ JNP (StatCan) ■ Nul-Faible ■ Modéré-Élevé ■ NP (Auto)



Catégories de joueurs selon la résidence (%)



Bref,

- NP (Auto) vs. JNP (StatCan)
Plus âgés
Plus d'hommes (Moins de femmes)
Moins Emploi+Études; Plus Ni emploi+Ni études
Plus Mariés/union libre; Moins Célibataires
- NP (Auto) vs. Nul/Faible
Profil similaire
- NP (Auto) vs. Modéré/Élevé
Plusieurs différences ...

Problème ...

On ne peut pas utiliser des variables associées aux problèmes de jeu pour imputer des valeurs probables aux données manquantes à l'IGPJ

...

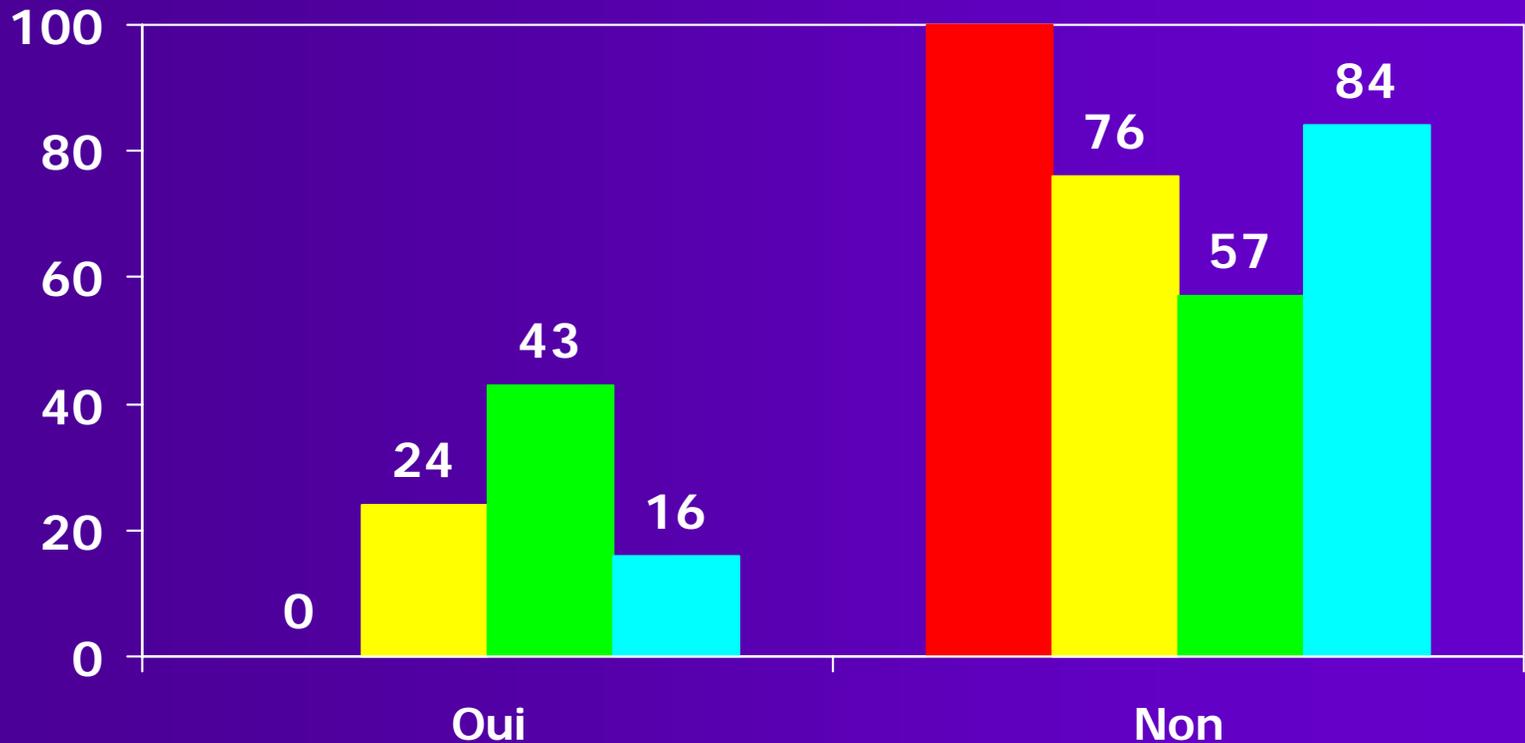
et ensuite étudier l'influence de ces variables sur le risque d'avoir un problème de jeu

Stratégie adoptée

- Hypothèse : Le risque augmente avec la dose
- Méthode
Modèle prédictif basé sur les activités de jeu
Échantillon: Répondants qui ont répondu à l'IGPJ
Variable dépendante : IGPJ
Variables indépendantes: Fréquence du jeu
- Sélection du modèle qui maximise la classification correcte des répondants à l'IGPJ
- Application du modèle aux NP(Auto)

Catégories de joueurs selon qu'ils jouent ou non 2 à 6 fois (ou plus) par semaine (%)

■ JNP (StatCan) ■ Nul-Faible ■ Modéré-Élevé ■ NP (Auto)



Prédiction de l' IGPJ à partir des 13 activités de jeu (n = 11 183 répondants à l'IGPJ)

- Régression linéaire - Méthode Stepwise (19% var.)

$$\begin{aligned} \text{IGPJ} = & .005 + .547 \text{ Vidéo hors casino} \\ & + .393 \text{ Autre jeu casino} \\ & + .160 \text{ Vidéo au casino} \\ & + .112 \text{ Bingo} \\ & + .133 \text{ Loterie sportive} \\ & + .054 \text{ Loterie instantanée} \\ & + .113 \text{ Courses} \\ & + .081 \text{ Investissement} \\ & - .019 \text{ Loterie type 6-49} \end{aligned}$$

Comparaison des valeurs prédites vs. observées

Valeurs observées	Valeurs prédites			
	JNP	Risque faible	Risque modéré	Risque élevé
JNP	6982	2271	172	0
Risque faible	347	545	124	0
Risque modéré	106	298	138	0
Risque élevé	21	90	88	1

Performance du modèle prédictif de l'IGPJ par catégorie de joueurs

	Sous-estimation	Correcte	Sur-estimation
JNP (StatCan)	-	74,0%	26,0%
Risque faible	34,2%	53,6%	12,2%
Risque modéré	74,5%	25,5%	-
Risque élevé	99,5%	0,5%	-

Distribution (%) des catégories de joueurs selon le traitement des données manquantes

	Sans imputation	Imputation Statcan	Imputation Régression
Non Joueurs	24,6	24,6	24,6
JNP (StatCan)	46,4	70,4	65,4
Risque faible	2,7	2,7	7,5
Risque modéré	1,5	1,5	1,6
Risque élevé	0,5	0,5	0,5
Non parieurs (auto-déclarés)	24,3	-	-

Association de IGPJ avec les facteurs classiques (B régression multiple excluant les non joueurs)

	Sans imputation	Imputation Statcan	Imputation Régression
constante	.495	.580	.667
Age	.046	.047	.032
Sexe	.075	.058	.068
Emploi/Études	-.028 (ns)	-.041	-.045
Emploi seul.	.155	.121	.127
Études seul.	-.234	-.242	-.279
Détresse psy.	.011	.008	.008

Alors, qu'est-ce qu'on a fait?

- Scénario # 2 Imputation StatCan
Non parieurs auto-déclarés = classés JNP

Parce que ..

Comparabilité des données

Maximise taille de l'échantillon

Modèle prédictif explique 19% de la variance IGPJ

Classification correcte faible pour joueurs à risque modéré/élevé

MAIS ...

Classification erronée potentielle ($\approx 5\%$)

Impact sur les résultats des analyses