

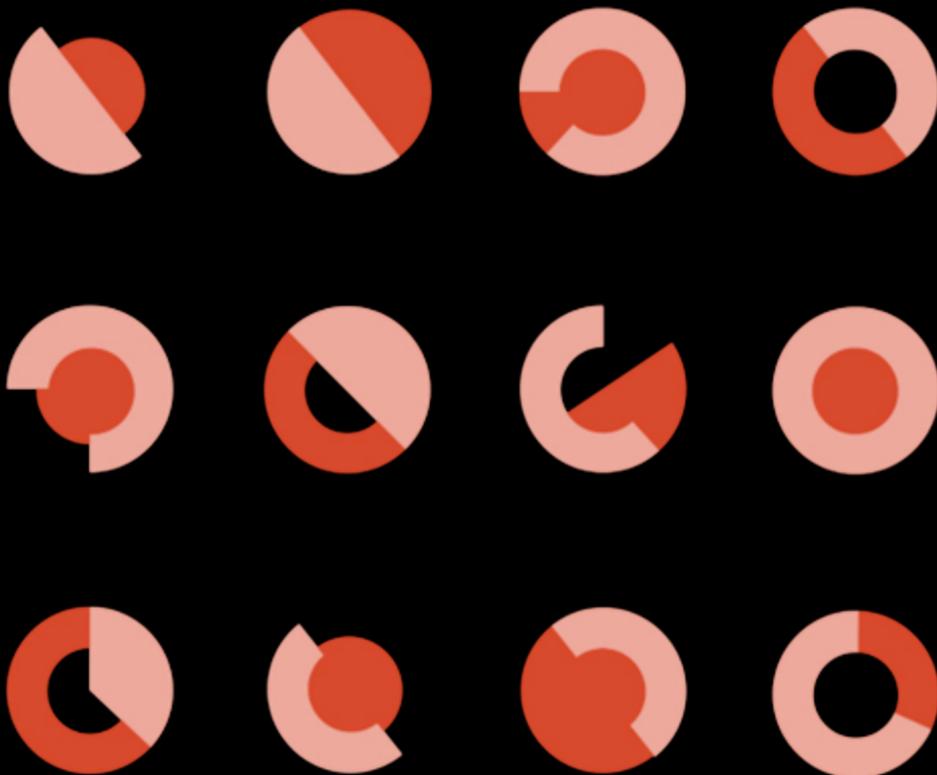


# LE DÉVELOPPEMENT DE MÉTHODES STATISTIQUES DE POINTE DANS LE DOMAINE DU JEU VIDÉO, DU RÊVE À LA RÉALITÉ

**Gabriel Boisvert-Beaudry**



**UBISOFT  
DATA OFFICE**



# LE COMMENCEMENT



# MON PARCOURS ACADÉMIQUE



Université   
de Montréal

Stages de recherche



Bacc en statistiques

Université   
de Montréal

M. Sc en statistiques

Stage en entreprise



MALA with annealed proposals  
A generalization of locally- and globally-  
balanced proposal distributions  
Gabriel Boisvert-Beaudry · Mylène Bédard

# MON PARCOURS PROFESSIONNEL



Stage  
Science des  
données



Analyste de  
données

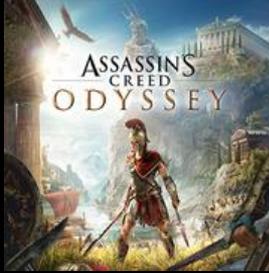


Hey, ça  
va?

Scientifique de  
données

# UBISOFT

~ 20 000 EMPLOYÉS DONT ~4000 À MONTRÉAL



# NOTRE ÉQUIPE

## Équipe analytique de jeux



Scientifiques de  
données



Analyste de  
données



## Clients

Designers de jeu

Experts  
monétisation

Architectes  
technique

# EXEMPLES DE QUESTIONS

QU'EST-CE QUI FAIT QUE LES JOUEURS  
ARRÊTENT DE JOUER?

EST-CE QUE L'EXPÉRIENCE EST TROP RUDE POUR LES  
NOUVEAUX JOUEURS?

QUELS SONT LES ARCHÉTYPES DE JOUEURS?

# LE PROBLÈME



# ÉVALUER L'IMPACT DU CONTENU DANS LE JEU



Contenu optionnel limité  
durant 1 semaine

Ce contenu en vaut-il la  
chandelle?

~~Cela a-t-il augmenté  
l'intérêt des joueurs  
pour le jeu?~~

Est-ce que leur temps de  
jeu a augmenté?



# EXPÉRIENCE RANDOMISÉE?

Nous sommes des experts!



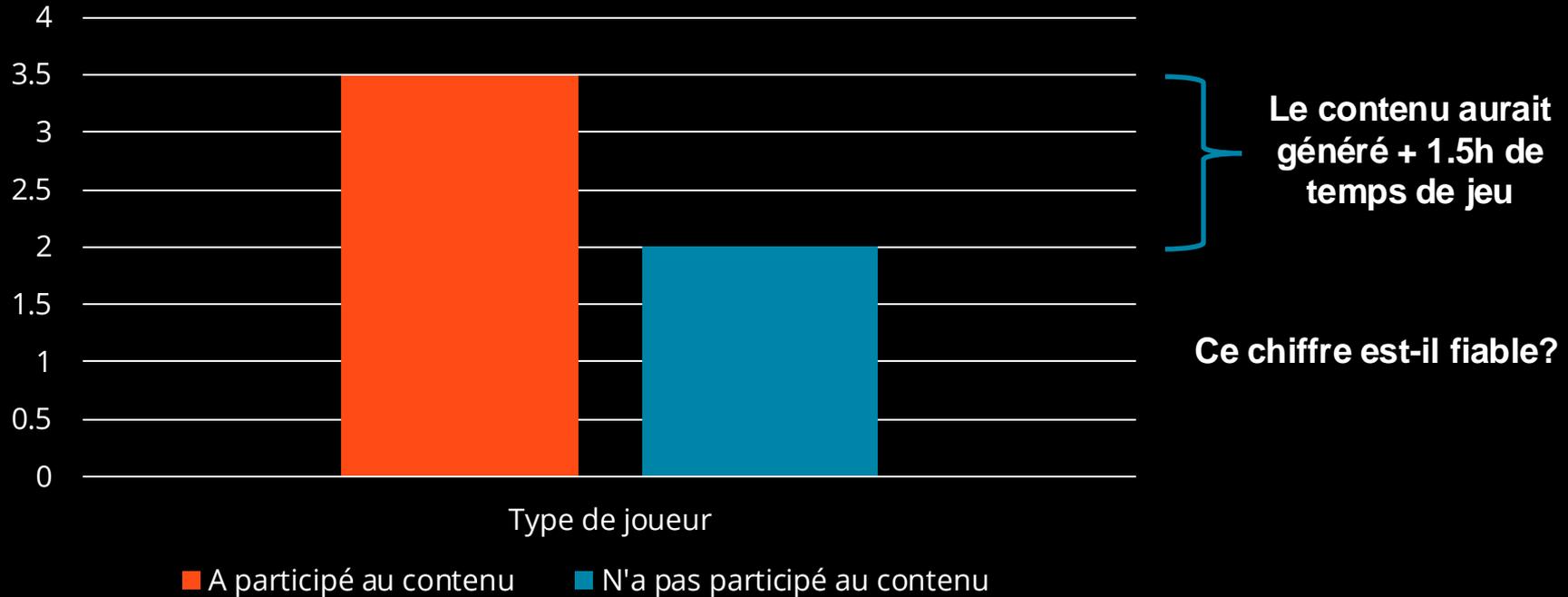
Contrôle: n'a pas le contenu

Traitement: a le contenu



# ON COMPARE LES 2 GROUPEES?

Temps de jeu moyen (h) la semaine suivante



# PEUT-ON COMPARER LES 2 GROUPEES?

Participants



Non Participants



vs

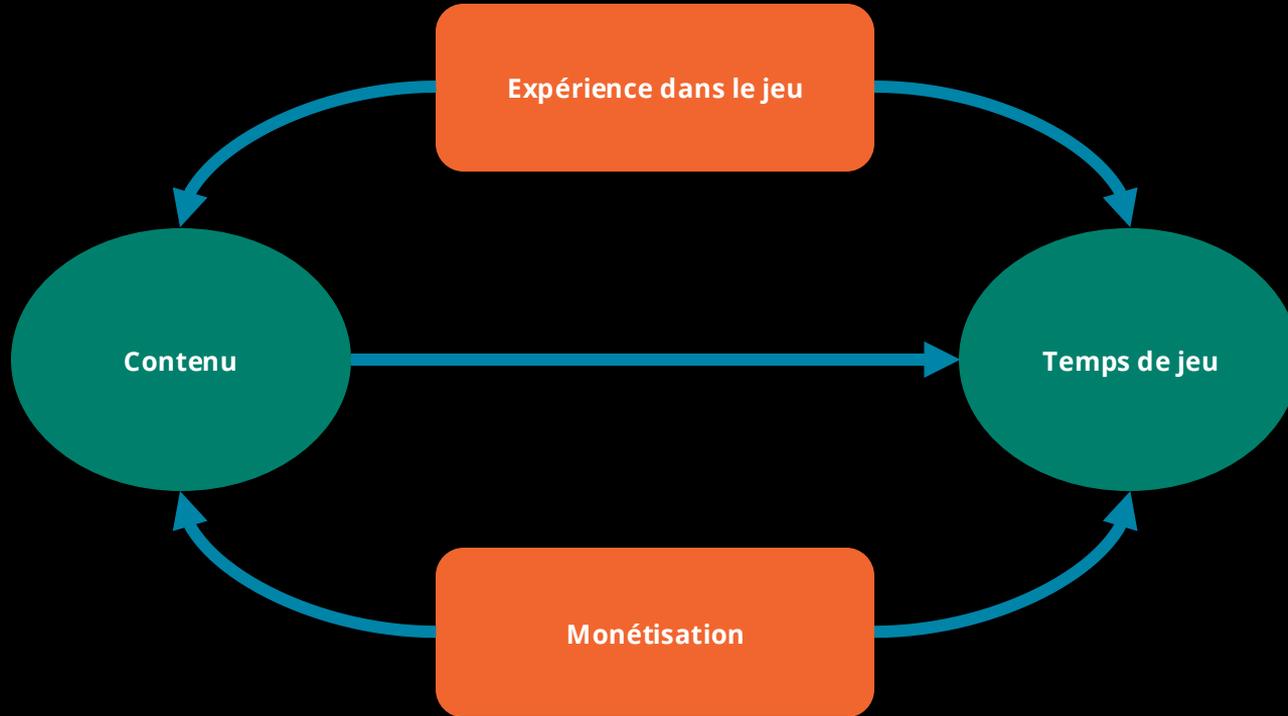
Les groupes ne  
sont pas  
randomisés



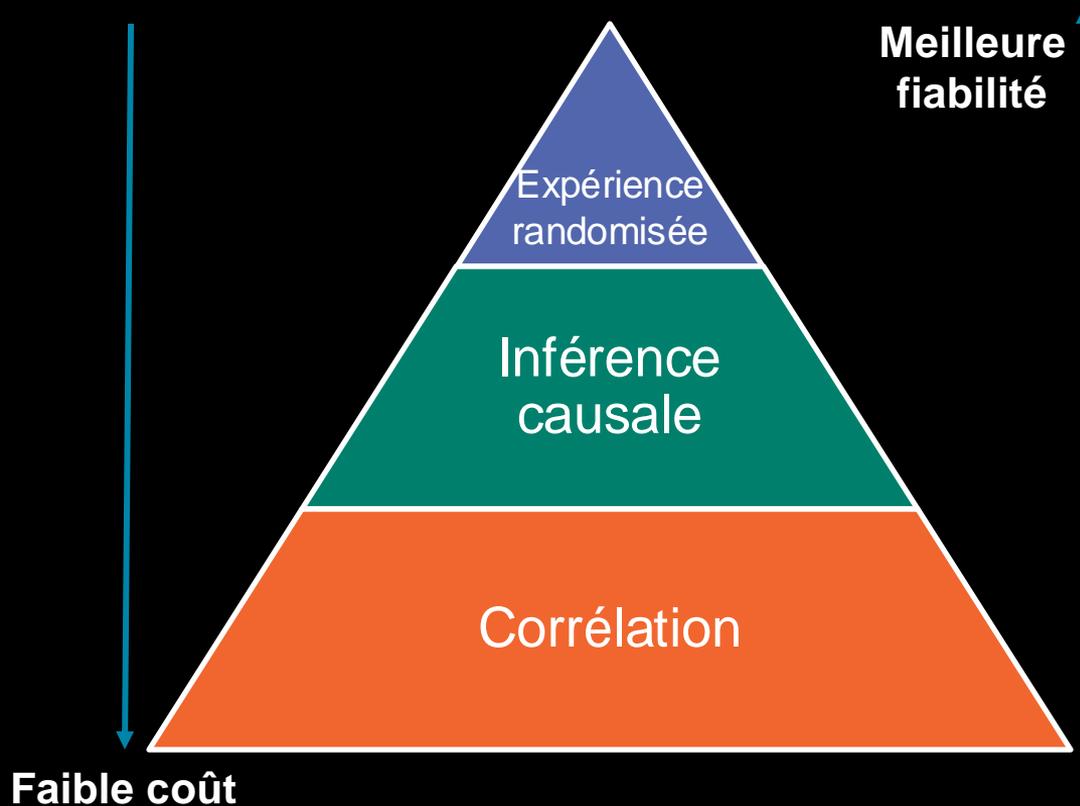
Certains joueurs ont plus d'expérience dans le jeu

D'autres ont monétisé dans le jeu

# LE PROBLÈME DES CONFONDEURS



# QUELLE EST LA SOLUTION?



Il existe une approche intermédiaire, l'**inférence causale**.

# LA SOLUTION



# INFÉRENCE CAUSALE

Ensemble de techniques permettant d'identifier des liens de cause à effet dans des études observationnelles

JOURNAL ARTICLE

## Smoking and Lung Cancer: Recent Evidence and a Discussion of Some Questions [Get access >](#)

Jerome Cornfield, William Haenszel, E. Cuyler Hammond, Abraham M. Lilienfeld, Michael B. Shimkin, Ernst L. Wynder

*JNCI: Journal of the National Cancer Institute*, Volume 22, Issue 1, January 1959, Pages 173–203, <https://doi.org/10.1093/jnci/22.1.173>

**Published:** 01 January 1959 [Article history](#) ▼

## The Environment and Disease: Association or Causation?

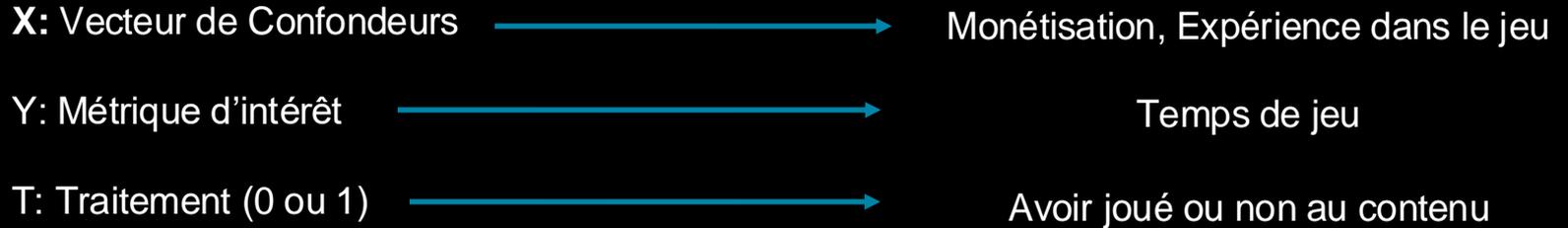
by Sir Austin Bradford Hill CBE DSC FRCP(hon) FRS  
(Professor Emeritus of Medical Statistics,  
University of London)

## The effectiveness of safety belts in preventing fatalities

[Leonard Evans](#)



# ADAPTATION AU JEU VIDÉO



## Modèle des contrefaits

$Y(0)$ : Temps de jeu qu'on aurait eu sans traitement

$Y(1)$ : Temps de jeu qu'on aurait eu avec traitement

On n'observe jamais  $Y(0)$  et  $Y(1)$  à la fois sur chaque personne.

# ESTIMATION

## Étude randomisée

On veut estimer

$$\text{Average treatment effect (ATE)} = E[Y(1)] - E[Y(0)]$$

On observe seulement  $Y|T=1$  et  $Y|T=0$ , le temps joué chez les traités vs non-traités. Est-ce que

$$E[Y(1)] - E[Y(0)] = E[Y|T = 1] - E[Y|T = 0] ?$$

Oui, si l'impact du traitement n'est pas affecté par le fait d'être traité ou non, car

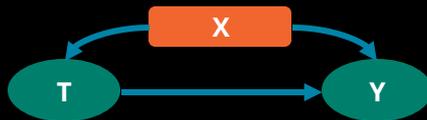
(Échangeabilité)

$$P(Y(1)) = P(Y(1)|T = 1) = P(Y(1)|T = 0)$$

# ESTIMATION

## Étude observationnelle

Les confondeurs  $X$  affectant  $T$  et  $Y$ , l'effet mesuré dans un groupe n'est plus indépendant du traitement!



Pour prendre  $X$  en compte, on utilise la méthode de **Inverse Probability Weighting**:

$$\widehat{ATE}_{IPW} = \frac{1}{n_T} \sum_{i=1}^n \frac{Y_i \mathbf{1}\{T_i = 1\}}{P(T_i = 1|X_i)} - \frac{1}{n_C} \sum_{i=1}^n \frac{Y_i \mathbf{1}\{T_i = 0\}}{1 - P(T_i = 1|X_i)}$$

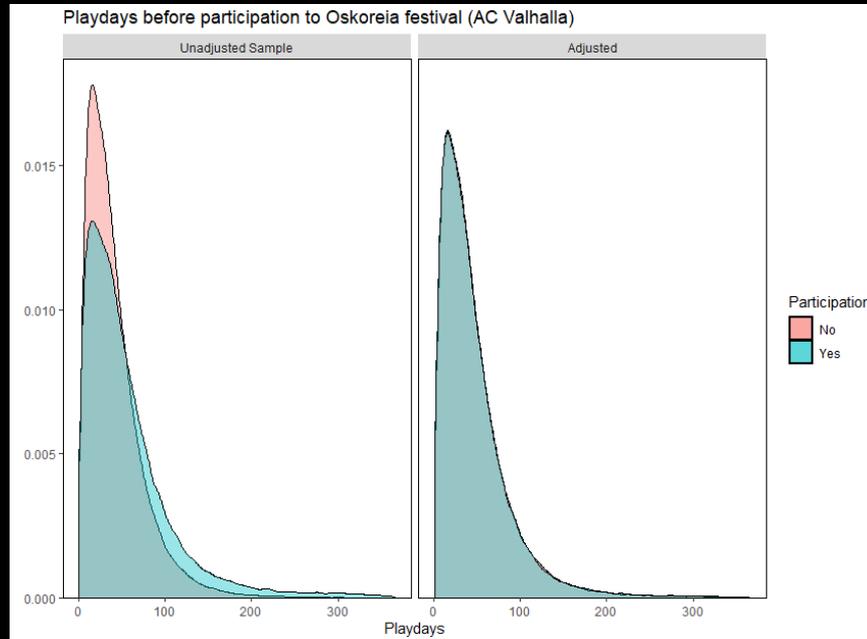
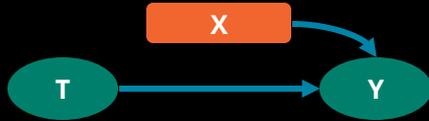
où  $P(T_i|X_i)$  est le **score de propension**.

On met plus de poids sur les gens ayant reçu le traitement, mais qui ressemblent au groupe contrôle, et vice-versa.

# ESTIMATION

## Effet des scores de propension

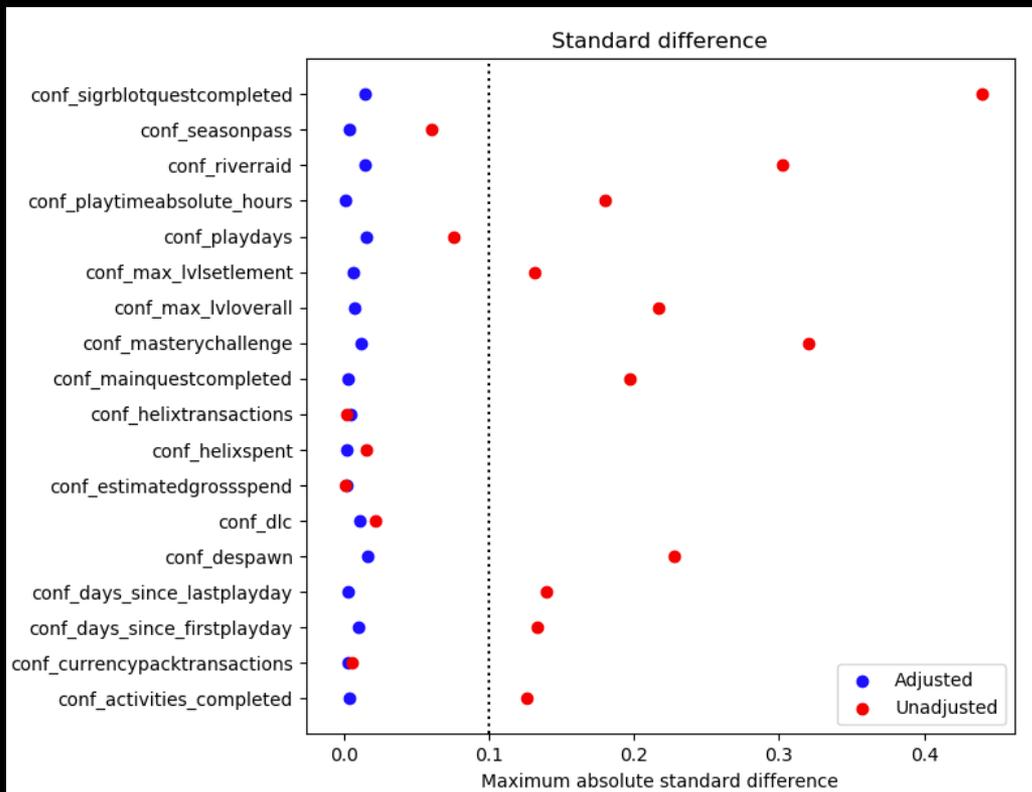
Dans cette population pondérée, la distribution des confondeurs est la même dans les deux groupes:



$P(T|X)$  doit être estimée

↓  
ML!

# COMMENT VALIDER LA MÉTHODE?

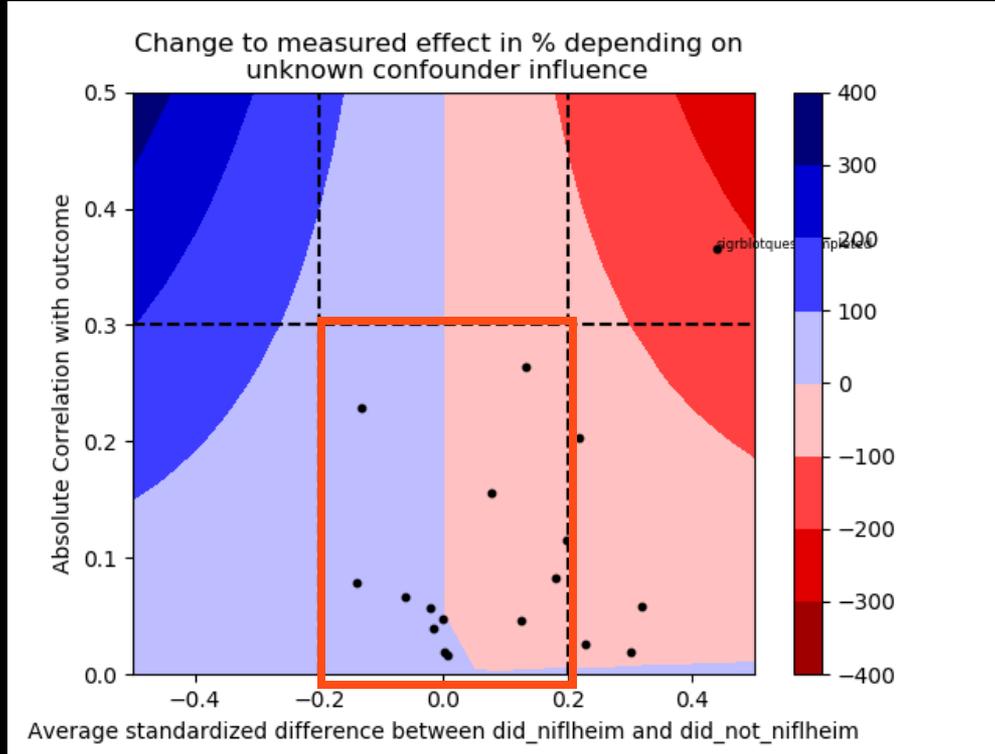


$P(T|X)$  n'a pas à être exacte

Doit juste balancer les confondeurs



# ET LES CONFONDEURS MANQUANTS?



On simule des confondeurs manquants et on regarde l'effet sur l'estimation

Y a-t-il de "faibles" confondeurs affectant beaucoup l'effet mesuré?

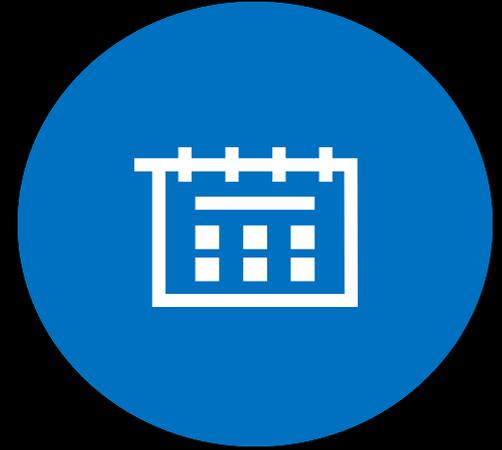
# DÉFIS D'ADAPTATION



**BIEN IDENTIFIER LE  
TRAITEMENT**



**LA CAUSE DOIT PRÉCÉDER  
L'EFFET**



**DÉFINIR SES PÉRIODES DE  
MESURES**

# L'IMPLANTATION



# ADOPTER UNE MÉTHODOLOGIE

## Équipe analytique de jeux



Scientifiques de données

Nous avons cette nouvelle méthode!



Analyste de données

Ça a l'air compliqué!

Est-ce fiable?

Peut-on l'avoir dans nos outils?

# ÉTAPE 1: VULGARISER

Branding



## SE FAIRE ENTENDRE

Participer aux présentations à l'interne



## DOCUMENTER

Avoir des tutoriels à portée de main aide à convaincre

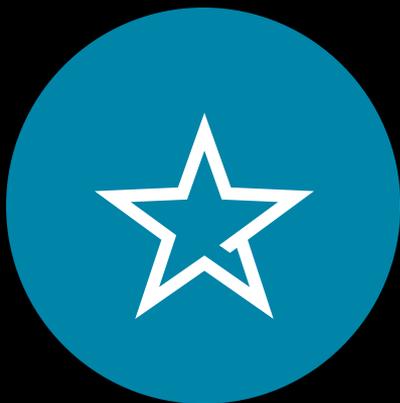
# ÉTAPE 2: UTILISER LES BONS OUTILS



## Développement de packages

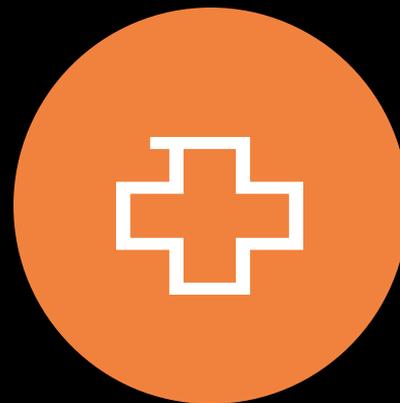
```
[2]: from udo_can_ds.ReleaseImpactMeter.main import ReleaseImpactMeter  
  
# Import PySpark module  
from pyspark.sql import SparkSession  
import matplotlib.pyplot as plt  
  
# initiating a spark session  
spark = SparkSession.builder.enableHiveSupport().getOrCreate()
```

# ÉTAPE 3: SÉCURISER LES MÉTHODES



## TROUVER DES CHAMPION.NES

Les accompagner dans le  
processus



## RÉPONDRE AUX INQUIÉTUDES

Analyse de sensibilité par exemple  
Simulations vs AB test

# MOTS DE LA FIN

## AIGUISEZ VOS TALENTS D'ORATEURS

Développer des méthodes statistiques de pointes est possible, mais cela demande de la patience et de l'endurance pour briser le statut quo.

## ADAPTEZ VOUS

Le monde technologique est vaste et une minorité a une formation en statistiques. Adaptez votre discours et vos méthodes à votre auditoire.

## SOYEZ CURIEUX

Notre formation ne fait pas en sorte que nous connaissons tout de la statistique. Restez informés et alertes des nouvelles tendances afin de les reconnaître dans votre travail.



# RESSOURCES

*What if, Robins, Hernan*

*Design of Observational Studies, Rosenbaum*





**MERCI!**

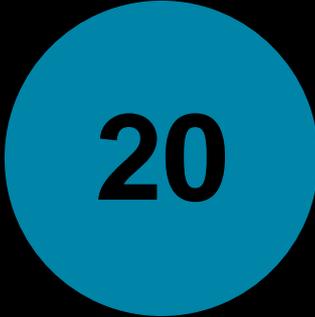
QUESTIONS?



# APPENDIX



# FALSE POSITIVES FOR CHEAT DETECTION



20

Unbans over  
84 days on  
6286 bans



# SKILL, MMR AND RANK POINTS

- **SKILL**

- Player skill determined by average skill ( $\mu$ ) and uncertainty about it ( $\sigma$ )

- **MMR**

- Match Making Rating used when sending request to find match. On R6S, usually  $MMR = \mu * 100$ .
- Some adjustments are added on the Standard Playlist to help newcomers

- **RANK POINTS**

- Points in Ranked displayed to Player, used to compute player's rank. MMR is used to compute Rank Points (could have also used skill instead without any issue)



# SQUAD AND TEAM

- **SQUAD**
  - All players that group up before launching match making.
  - We consider a solo player to be in a squad of 1
  
- **TEAM**
  - A ranked match is composed of two teams
  - Each team is composed of 1-5 squads



# SKILL RATING SIMULATION

How to determine  $\mu_{\text{phantom}}$ ?

- Compute skill rating of players based on match history
- Add phantom players of a given dimensionality and granularity
  - Squad size (1-5)
  - Team average skill (10-22, 22-34, ...)
  - Team skill disparity (0-1, ...)
- Apply skill rating update after each match on players and phantom players

Squad size

	1	2	3	4	5
10-22	0	1.9	1.6	1.8	1.8
22-25.2	0.1	1.9	1.7	2.0	2.2
25.2-27.9	0.3	2.4	2.3	2.5	3.0
27.9-31.7	0.7	3.5	3.2	3.4	3.8
31.7+	0.8	4.5	4.4	4.6	5.0

Team skill disparity

	0-1	1-1.55	1.55-2.1	2.1-2.7	2.7+
10-22	0.1	1.0	1.0	0.4	0.2
22-25.2	0	1.0	1.4	1.0	1.9
25.2-27.9	0.05	1.4	2.0	2.0	3.0
27.9-31.7	1.9	2.3	2.4	2.2	3.2
31.7+	2.4	3.1	3.2	3.4	4.2

Skill bracket



# EXAMPLE

Adding phantom player in match making or skill updates



Skill bracket

Squad size

	1	2	3	4	5
10-22	0	1.9	1.6	1.8	1.8
22-25.2	0.1	1.9	1.7	2.0	2.2
25.2-27.9	0.3	2.4	2.3	2.5	3.0
27.9-31.7	0.7	3.5	3.2	3.4	3.8
31.7+	0.8	4.5	4.4	4.6	5.0

Updated team skill mean:

$$\mu_A + \mu_{\text{phantom}} = (27.3 * 2 + 28.4 * 3 + 2.4 * 2 + 3.2 * 3) / 5 = 30.8$$

$$\mu_A = (27.3 * 2 + 28.4 * 3) / 5 = 28.0$$

