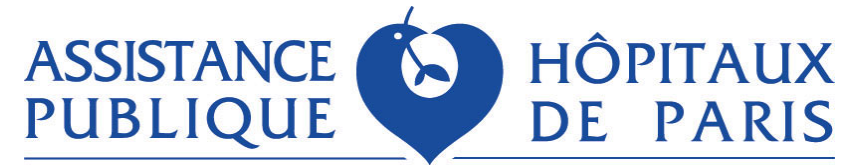


SOUTENANCE D'ALTERNANCE

ALTERNANCE AU CEREDIH (HÔPITAL NECKER, AP-HP)

Merlin Simsen Baratault - BUT3 SD EMS



INTRODUCTION

- Alternance de 2 ans au CEREDIH
- Encadrants : Mickaël Alligon, Nizar Mahlaoui, François Montestruc
- Objectif : présenter mes **missions**, mes **apports**, les **compétences** développées, et apports à mon projet professionnel

LE CEREDIH

- Créée en 2005 - Plan maladies rares
- Registre national **> 10 000 patients**
- **Réseau français & européen (ESID)**
- Collecte, analyse, valorisation
- Équipes : ARC, Docteurs, Data
- Mon rôle : alternant Data

SERVICE DATA

- 2 personnes : Mickaël Alligon (Datamanager / Data analyst), moi (alternant)
- Rôle : gestion des données, analyses statistiques, valorisation (rapports, articles)
- Outils : R, R Markdown, Excel, PowerPoint

APERÇU DE MES MISSIONS

- **Refonte du Data Management** (optimisation, robustesse, lisibilité)
- **Modernisation** du rapport d'activité
- Rédaction d'un **article méthodologique** sur le modèle de **Cox**

MISSION 1 - REFONTE DU DATA MANAGEMENT

CONTEXTE & OBJECTIFS

- Scripts historiques (~9000 lignes), compilation ~2h, erreurs fréquentes
- Objectifs : **rapidité, robustesse, lisibilité**

MISSION 1 - AVANT

Analyse du script existant → remplacer boucles for()

```
1 # Boucle lente et fragile
2 for (i in 1:nrow(Esid))
3 {
4   x < rep(NA, length(IG_date))
5   for (j in 1:length(IG_date))
6   {
7     if (as.character(Esid[i, IG_status[j]]) %in% "Y")
8     {
9       Esid$date_first_IG_replacement[i] <- date_ceredih(Esid[i, IG_date[j]])
10      break()
11    }
12  }
13 }
```

Exemple de code: passages ligne par ligne, boucles imbriquées

MISSION 1 - AVANT

Analyse du script existant → remplacer boucles for()

```
1 # Boucle lente et fragile
2 for (i in 1:nrow(Esid))
3 {
4   x < rep(NA, length(IG_date))
5   for (j in 1:length(IG_date))
6   {
7     if (as.character(Esid[i, IG_status[j]]) %in% "Y")
8     {
9       Esid$date_first_IG_replacement[i] <- date_ceredih(Esid[i, IG_date[j]])
10      break()
11    }
12  }
13 }
```

Exemple de code: passages ligne par ligne, boucles imbriquées

MISSION 1 - AVANT

Analyse du script existant → remplacer boucles for()

```
1 # Boucle lente et fragile
2 for (i in 1:nrow(Esid))
3 {
4   x < rep(NA, length(IG_date))
5   for (j in 1:length(IG_date))
6   {
7     if (as.character(Esid[i, IG_status[j]]) %in% "Y")
8     {
9       Esid$date_first_IG_replacement[i] <- date_ceredih(Esid[i, IG_date[j]])
10      break()
11    }
12  }
13 }
```

Exemple de code: passages ligne par ligne, boucles imbriquées

MISSION 1 - AVANT

Analyse du script existant → remplacer boucles for()

```
1 # Boucle lente et fragile
2 for (i in 1:nrow(Esid))
3 {
4   x < rep(NA, length(IG_date))
5   for (j in 1:length(IG_date))
6   {
7     if (as.character(Esid[i, IG_status[j]]) %in% "Y")
8     {
9       Esid$date_first_IG_replacement[i] <- date_ceredih(Esid[i, IG_date[j]])
10      break()
11    }
12  }
13 }
```

Exemple de code: passages ligne par ligne, boucles imbriquées

MISSION 1 - AVANT

Analyse du script existant → remplacer boucles for()

```
1 # Boucle lente et fragile
2 for (i in 1:nrow(Esid))
3 {
4   x < rep(NA, length(IG_date))
5   for (j in 1:length(IG_date))
6   {
7     if (as.character(Esid[i, IG_status[j]]) %in% "Y")
8     {
9       Esid$date_first_IG_replacement[i] <- date_ceredih(Esid[i, IG_date[j]])
10      break()
11    }
12  }
13 }
```

Exemple de code: passages ligne par ligne, boucles imbriquées

MISSION 1 - APRÈS

```
1 # Vectorisation claire et robuste
2 Esid$date_first_IG_replacement <- mapply(extract_element,
3     which(is.na(Esid$first_ig_replacement)),
4     Esid$first_ig_replacement[which(is.na(Esid$first_ig_replacement))],
5     MoreArgs = list(base = Esid[, IG_date]))
```

- Temps total réduit : ~2h → ~15 min
- Processus plus **fiable & maintenable**
- Gain de **temps**

MISSION 1 - APRÈS

```
1 # Vectorisation claire et robuste
2 Esid$date_first_IG_replacement <- mapply(extract_element,
3     which(is.na(Esid$first_ig_replacement)),
4     Esid$first_ig_replacement[which(is.na(Esid$first_ig_replacement))],
5     MoreArgs = list(base = Esid[, IG_date]))
```

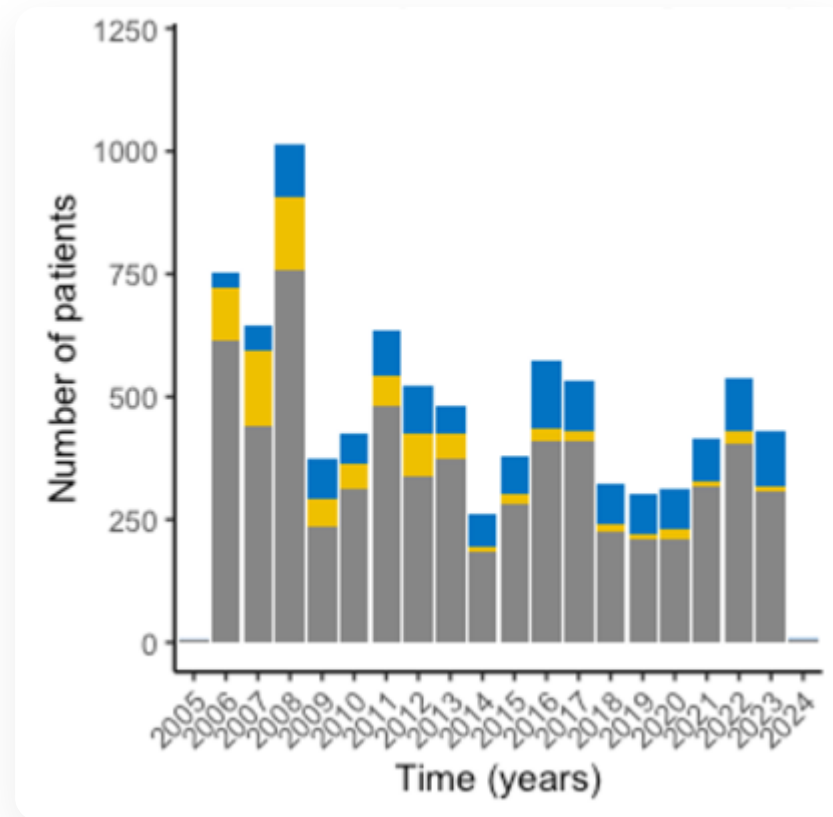
- Temps total réduit : ~2h → ~15 min
- Processus plus **fiable & maintenable**
- Gain de **temps**

MISSION 2 - RAPPORT D'ACTIVITÉ

CONTEXTE

- Peu lisible, graphiques figés, limites du format classique
- Objectifs : moderniser, rendre **interactif**, mieux valoriser les données

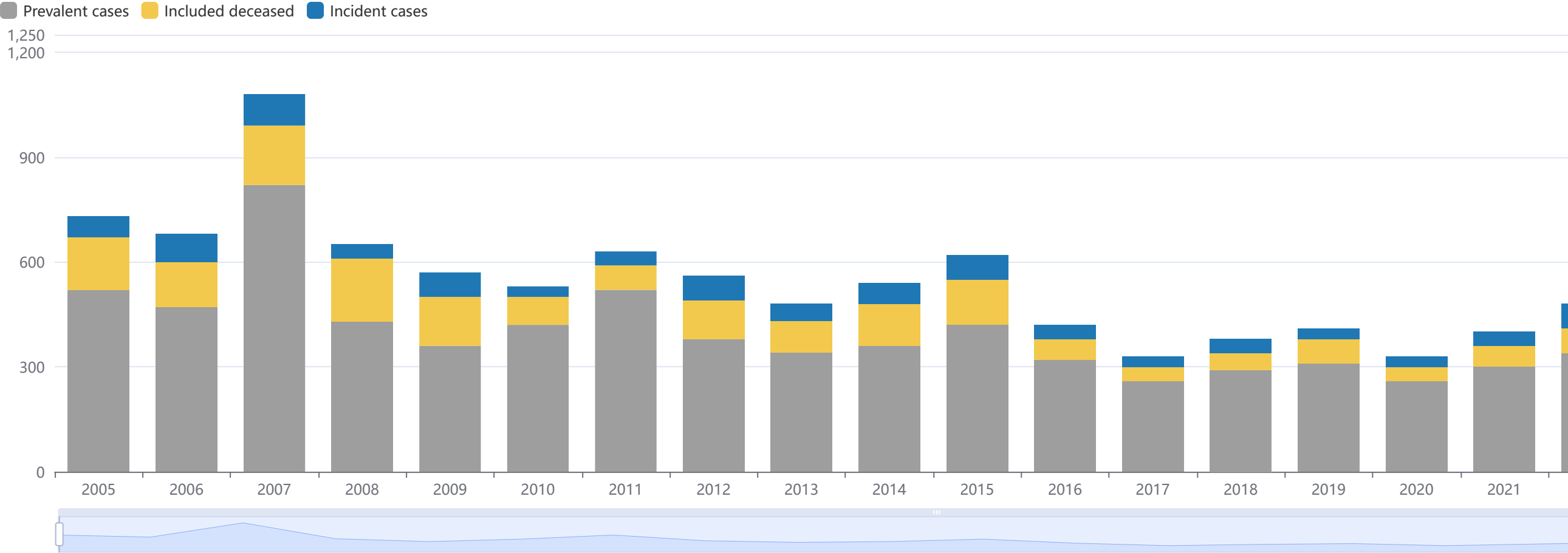
AVANT



graphique présentant le nombre d'inclusion par année

APRÈS

Nombre de patients par année



MISSION 2 - RÉSULTATS

- Rapport **interactif** (≈ 70 pages) - utilisé en congrès
- Newsletter modernisée

MISSION 3 - ARTICLE SUR COX

Article méthodologique sur le modèle de Cox (guide pratique)

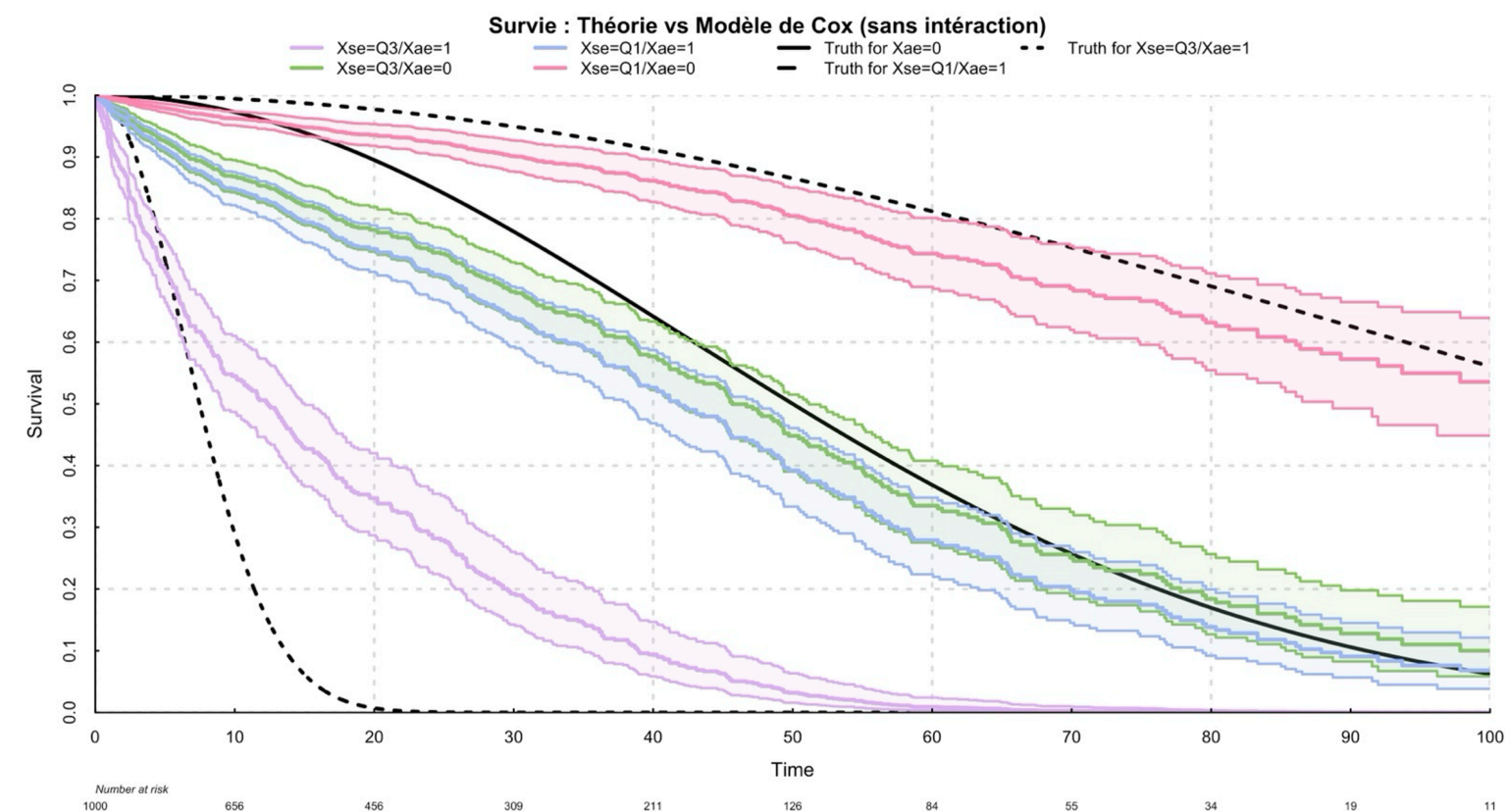
- Suite du papier sur *Kaplan-Meier*
- Public: cliniciens non-statisticiens
- Objectif : expliquer les biais méthodologiques fréquents

MÉTHODE

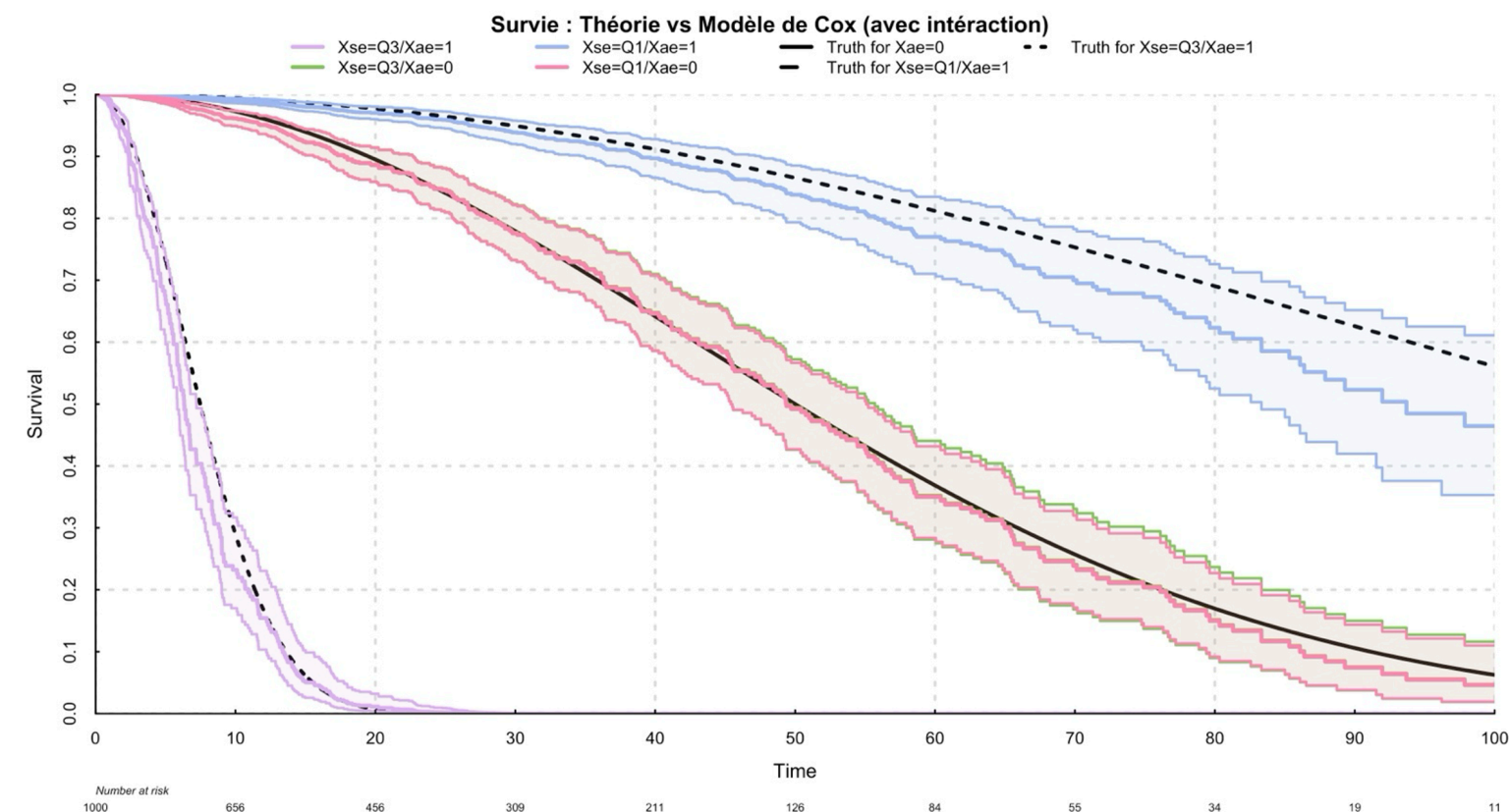
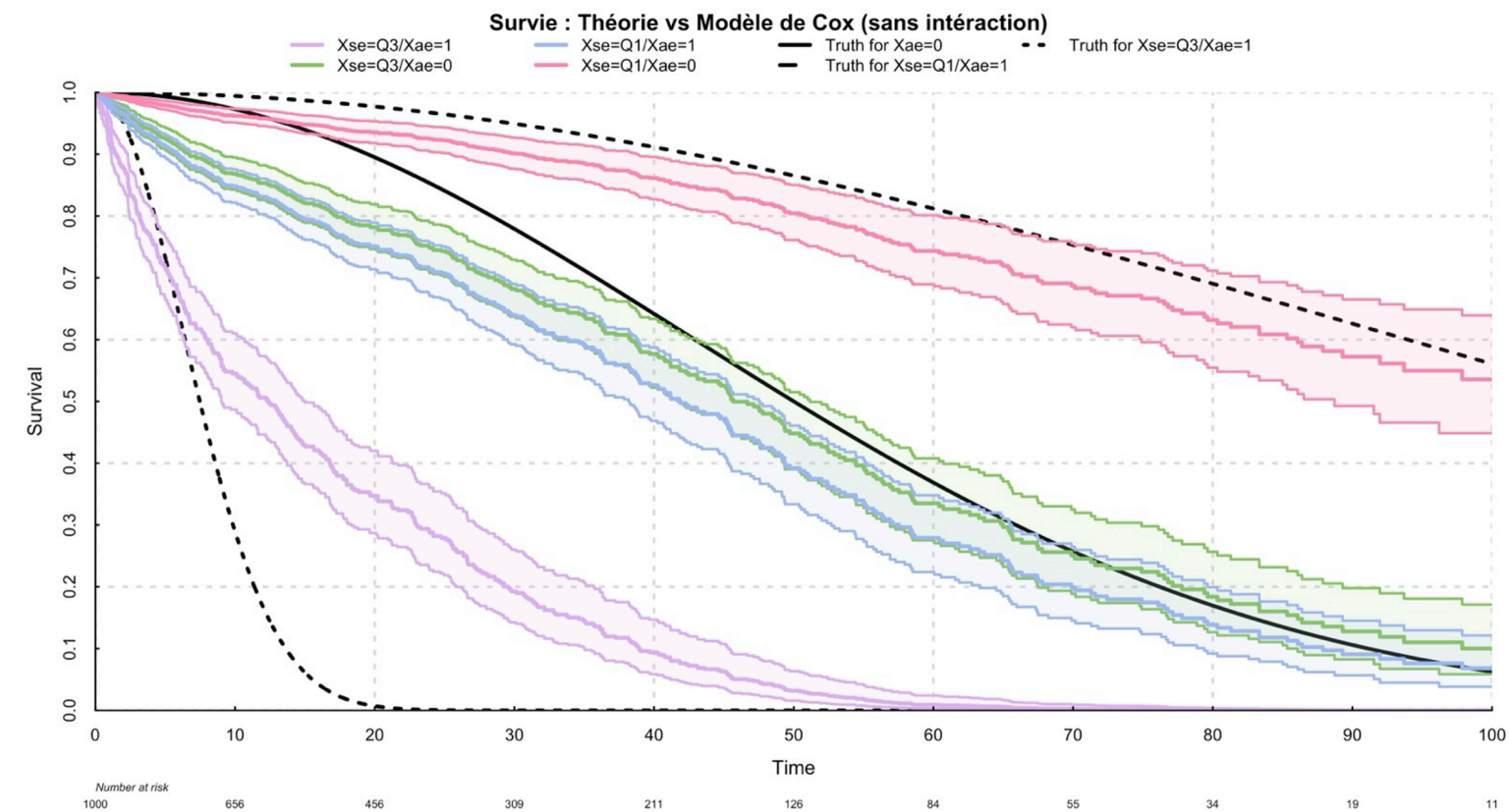
- Simulation données
- Covariables: binaire (effet réel) + continue (sans effet)
- Introduction des biais: troncature, censure, risques compétitifs, événements récurrents, interactions, non-proportionnalité
- Comparaison modèles **naïfs vs corrigés**

Exemple sur l'interaction, modèle naïf biaisé vs modèle corrigé

Exemple sur l'interaction, modèle naïf biaisé vs modèle corrigé



Exemple sur l'interaction, modèle naïf biaisé vs modèle corrigé



MISSION 3 - RÉSULTATS

- Article en cours de finalisation
- Outil pédagogique pour cliniciens - open source
- simulation de données complexe, survie, rédaction scientifique, gestion de projet

PORTFOLIO - COMPÉTENCES DU BUT SD

- **Traiter des données**

Refonte du Data Management, automatisation, qualité des données

- **Analyser statistiquement**

Analyses avancées, article sur le modèle de Cox, simulations

- **Valoriser une production**

Rapport d'activité interactif, newsletter modernisée

- **Modéliser les données**

Modèles de survie, gestion des biais, visualisations pédagogiques

COMPÉTENCES TRANSVERSALES

- Autoformation (survie, simulation, visualisation)
- Gestion de projet (planification, suivi, évaluation)
- Communication interdisciplinaire (clinique ↔ data)
- Travail en équipe et rigueur scientifique

BILAN & PERSPECTIVES

- **Techniques** : R, data management, analyse de survie, dataviz interactive
- **Professionnelles** : rigueur, vulgarisation scientifique, travail en équipe médicale
- **Impact personnel** : confirmation de mon orientation professionnelle

REMERCIEMENTS

Merci à toute l'équipe du CEREDIH pour leur accueil et leur accompagnement

Merci à mes encadrants : Mickaël Alligon, Nizar Mahlaoui, François Montestruc

Merci aux enseignants du BUT SD et à l'Université Paris Cité

Merci à vous pour votre attention !

DES QUESTIONS ?