
Modèles conjoints avec risques concurrents : analyse de données longitudinales de qualité de vie en oncologie en présence de différents types de sorties d'étude

Benjamin Cuer (doctorant)
Encadrement : Caroline Mollevi et Célia Touraine

`benjamin.cuer@icm.unicancer.fr`

Journées de « Statistiques et santé », Paris

11 octobre 2019

Mesure de la qualité de vie relative à la santé (QdV)

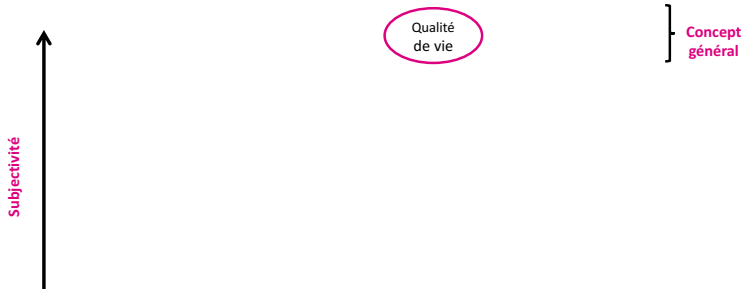
Qualité
de vie

} Concept
général

Une notion complexe

Évaluée à l'aide de questionnaires

Mesure de la qualité de vie relative à la santé (QdV)



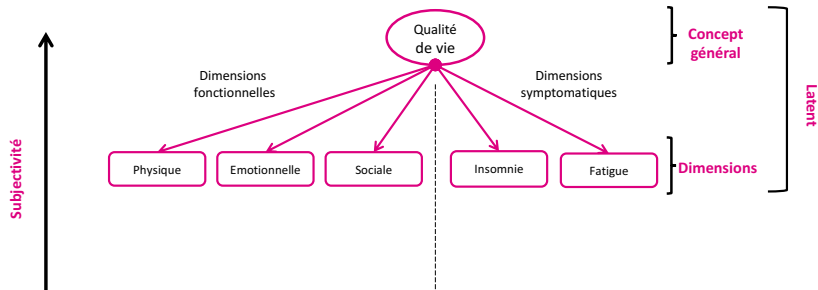
Une notion complexe

► Subjective

Évaluée à l'aide de questionnaires

► Auto-administrés

Mesure de la qualité de vie relative à la santé (QdV)



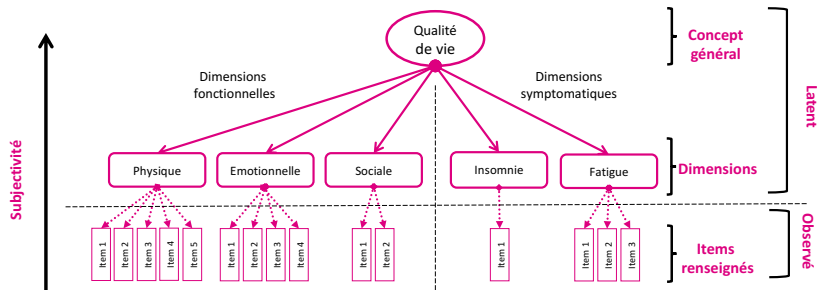
Une notion complexe

- Subjective
- Multidimensionnelle

Évaluée à l'aide de questionnaires

- Auto-administrés

Mesure de la qualité de vie relative à la santé (QdV)



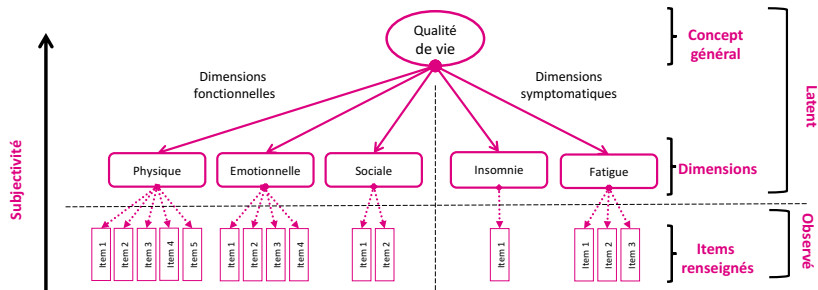
Une notion complexe

- Subjective
- Multidimensionnelle

Évaluée à l'aide de questionnaires

- Auto-administrés
- Items regroupés par dimension

Mesure de la qualité de vie relative à la santé (QdV)



Une notion complexe

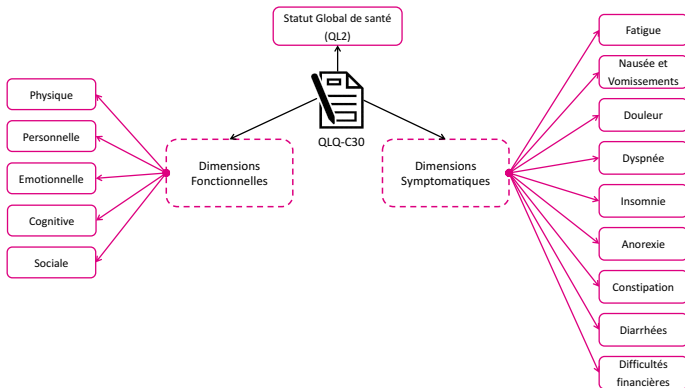
- Subjective
- Multidimensionnelle
- Dynamique

Évaluée à l'aide de questionnaires

- Auto-administrés
- Items regroupés par dimension
- Plusieurs évaluations au cours du temps

Questionnaire EORTC QLQ-C30

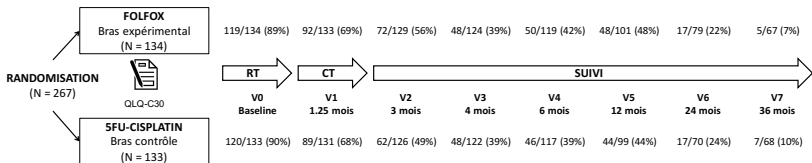
- 30 items regroupés en 15 échelles :



- Pour chaque échelle :
1 ou plusieurs variables qualitatives ordinales → **Score de 0 à 100**

Essai clinique PRODIGE5/ACCORD17

- ▶ Essai randomisé de phase II-III
- ▶ $N = 267$ patients avec un cancer de l'oesophage localement avancé traité par (radio)-chimiothérapie
 - FOLFOX : bras expérimental
 - 5FU-Cisplatine : bras contrôle
- ▶ **Critère principal** : PFS
 - Pas de différence entre les 2 bras de traitement¹
- ▶ **Critère secondaire** : QdV (mesurée par le QLQ-C30)
 - Compliance au questionnaire :



- Analyse de la QdV par des modèles linéaires mixtes²

1. Conroy et al., Lancet Oncology, 2014
2. Mollevi et al., EJC, 2017

Analyse des données longitudinales de QdV

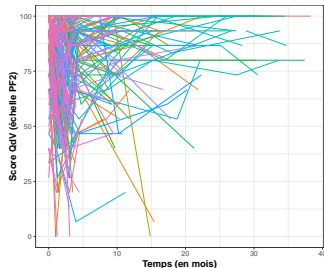
Modèle linéaire mixte

Le score de QdV du patient i au temps t s'écrit :

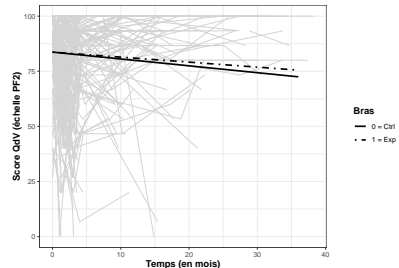
$$y_i(t) = \underbrace{\beta^T X_i(t)}_{\text{Trajectoire moyenne}} + \underbrace{b_i^T Z_i(t)}_{\text{Déviations individuelles}} + \varepsilon_i(t),$$

avec β le vecteur des effets fixes et b_i celui des effets aléatoires

Trajectoires individuelles observées



Trajectoires moyennes prédites



Données manquantes dans les analyses longitudinales

Types de données manquantes

- ▶ **MCAR/MAR** : le mécanisme de données manquantes est **indépendant** de la QdV non observée \Rightarrow Données manquantes non informatives
- ▶ **MNAR** : le mécanisme de données manquantes **peut dépendre** de la QdV non observée \Rightarrow Données manquantes informatives

Données manquantes dans les analyses longitudinales

Types de données manquantes

- **MCAR/MAR** : le mécanisme de données manquantes est **indépendant** de la QdV non observée \Rightarrow Données manquantes non informatives
- **MNAR** : le mécanisme de données manquantes **peut dépendre** de la QdV non observée \Rightarrow Données manquantes informatives

Sorties d'étude

- Données manquantes de structure monotone

Patient	Visite 1	Visite 2	Visite 3	Visite 4	Visite 5
1	83.3	100	66.7	66.7	×
2	91.7	25.0	25.0	×	×
3	66.7	66.7	×	×	×
4	50.0	×	×	×	×

Données manquantes dans les analyses longitudinales

Types de données manquantes

- ▶ **MCAR/MAR** : le mécanisme de données manquantes est **indépendant** de la QdV non observée \Rightarrow Données manquantes non informatives
- ▶ **MNAR** : le mécanisme de données manquantes **peut dépendre** de la QdV non observée \Rightarrow Données manquantes informatives

Sorties d'étude

- ▶ Données manquantes de structure monotone
- ▶ Non liées à la QdV (e.g. déménagement)
 \Rightarrow **non informatives**

Patient	Visite 1	Visite 2	Visite 3	Visite 4	Visite 5
1	83.3	100	66.7	66.7	×
2	91.7	25.0	25.0	×	×
3	66.7	66.7	×	×	×
4	50.0	×	×	×	×

Données manquantes dans les analyses longitudinales

Types de données manquantes

- ▶ **MCAR/MAR** : le mécanisme de données manquantes est **indépendant** de la QdV non observée \Rightarrow Données manquantes non informatives
- ▶ **MNAR** : le mécanisme de données manquantes **peut dépendre** de la QdV non observée \Rightarrow Données manquantes informatives

Sorties d'étude

- ▶ Données manquantes de structure monotone
- ▶ Non liées à la QdV (e.g. déménagement)
 \Rightarrow **non informatives**
- ▶ Liées à la QdV (e.g. sortie d'étude suite à un événement clinique)
 \Rightarrow **informatives**

Patient	Visite 1	Visite 2	Visite 3	Visite 4	Visite 5	Cause
1	83.3	100	66.7	66.7	×	
2	91.7	25.0	25.0	×	×	
3	66.7	66.7	×	×	×	Décès
4	50.0	×	×	×	×	Décès

Problématique et objectifs

Problématique

- ▶ Biais des estimations du modèle mixte si sorties d'étude informatives
- ▶ Prendre en compte l'association entre la QdV et les sorties d'étude :

Problématique et objectifs

Problématique

- Biais des estimations du modèle mixte si sorties d'étude informatives
- Prendre en compte l'association entre la QdV et les sorties d'étude :

⇒ Utilisation d'un **modèle conjoint**

= Modèle linéaire mixte + modèle de temps jusqu'à événement

Problématique et objectifs

Problématique

- Biais des estimations du modèle mixte si sorties d'étude informatives
- Prendre en compte l'association entre la QdV et les sorties d'étude :

⇒ Utilisation d'un **modèle conjoint**

= Modèle linéaire mixte + modèle de temps jusqu'à événement

- ↔ Sorties d'étude suite au décès = supposées plus informatives
- ↔ Distinguer **2 causes de sorties d'étude** ⇒ **Risques concurrents**

Problématique et objectifs

Problématique

- ▶ Biais des estimations du modèle mixte si sorties d'étude informatives
- ▶ Prendre en compte l'association entre la QdV et les sorties d'étude :

⇒ Utilisation d'un **modèle conjoint**

= Modèle linéaire mixte + modèle de temps jusqu'à événement

- ↪ Sorties d'étude suite au décès = supposées plus informatives
- ↪ Distinguer **2 causes de sorties d'étude** ⇒ **Risques concurrents**

Objectifs

- ▶ Evaluer la performance d'un modèle conjoint qui distingue les causes de sorties d'étude
 - Modèle 1 = conjoint simple (sans distinguer les causes de sorties d'étude)
 - Modèle 2 = conjoint avec risques concurrents (sortie d'étude liée ou non liée au décès)
1. Application aux données de l'essai clinique PRODIGE5/ACCORD17 pour analyser les données de QdV
 2. Etude de simulations pour comparer les 2 modèles

Modèles conjoints

Modélisation des données longitudinales du score de QdV

Le score de QdV du patient i au temps t s'écrit :

$$\begin{aligned} y_i(t) &= y_i^*(t) + \varepsilon_i(t) \\ &= \beta_0 + \beta_1 t + \beta_2 \{t \times \text{bras}_i\} + b_{0i} + b_{1i} t + \varepsilon_i(t) \end{aligned}$$

$$\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim \mathcal{N}(0, \Sigma) \text{ avec } \Sigma = \begin{pmatrix} \sigma_{b_0}^2 & \sigma_{b_0 b_1} \\ \sigma_{b_0 b_1} & \sigma_{b_1}^2 \end{pmatrix} \text{ et } \varepsilon_i(t) \sim \mathcal{N}(0, \sigma^2)$$

Modélisation du temps jusqu'à événement

Modèle de survie avec comme risque instantané pour $t \geq 0$:

$$h_i(t) = h_0(t) \exp\{\gamma \text{bras}_i + \alpha y_i^*(t)\} \quad (\text{M1})$$

Modèles conjoints

Modélisation des données longitudinales du score de QdV

Le score de QdV du patient i au temps t s'écrit :

$$\begin{aligned} y_i(t) &= y_i^*(t) + \varepsilon_i(t) \\ &= \beta_0 + \beta_1 t + \beta_2 \{t \times \text{bras}_i\} + b_{0i} + b_{1i} t + \varepsilon_i(t) \end{aligned}$$

$$\begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} \sim \mathcal{N}(0, \Sigma) \text{ avec } \Sigma = \begin{pmatrix} \sigma_{b_0}^2 & \sigma_{b_0 b_1} \\ \sigma_{b_0 b_1} & \sigma_{b_1}^2 \end{pmatrix} \text{ et } \varepsilon_i(t) \sim \mathcal{N}(0, \sigma^2)$$

Modélisation du temps jusqu'à événement

Modèle de survie avec comme risque instantané pour $t \geq 0$:

$$h_i(t) = h_0(t) \exp\{\gamma \text{bras}_i + \alpha y_i^*(t)\} \quad (\text{M1})$$

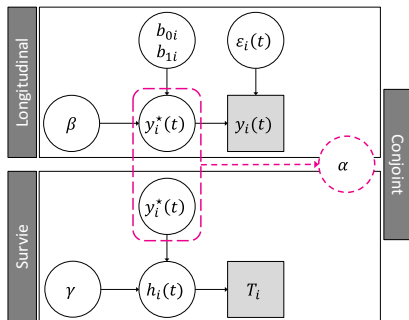
Modèle à 2 risques concurrents avec comme risques instantanés pour $t \geq 0$:

$$\begin{aligned} h_{i1}(t) &= h_{01}(t) \exp\{\gamma_1 \text{bras}_i + \alpha_1 y_i^*(t)\}, & \text{cause 1} \\ h_{i2}(t) &= h_{02}(t) \exp\{\gamma_2 \text{bras}_i + \alpha_2 y_i^*(t)\}, & \text{cause 2} \end{aligned} \quad (\text{M2})$$

Représentation et interprétation des modèles

Modèle conjoint (M1)

$$\begin{aligned}
 y_i(t) &= y_i^*(t) + \varepsilon_i(t) \\
 &= \beta_0 + \beta_1 t + \beta_2 \{t \times \text{bras}_i\} \\
 &\quad + b_{0i} + b_{1i} t + \varepsilon_i(t) \\
 h_i(t) &= h_0(t) \exp\{\gamma \text{bras}_i + \alpha y_i^*(t)\}
 \end{aligned}$$



- β_1 : **pente** = changement par unité de temps dans le bras contrôle
- β_2 : **effet d'interaction** = différence de pente entre le bras expérimental et le bras contrôle
- γ : **effet du bras** sur le risque de sortie d'étude
- α : degré d'**association** entre QdV et risque de sortie d'étude

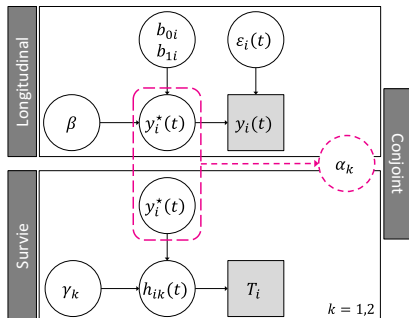
Représentation et interprétation des modèles

Modèle conjoint avec risques concurrents (M2)

$$y_i(t) = y_i^*(t) + \varepsilon_i(t)$$

$$= \beta_0 + \beta_1 t + \beta_2 \{t \times \text{bras}_i\} + b_{0i} + b_{1i}t + \varepsilon_i(t)$$

$$h_{ik}(t) = h_{0k}(t) \exp\{\gamma_k \text{bras}_i + \alpha_k y_i^*(t)\}$$



- β_1 : **pen**te = changement par unité de temps dans le bras contrôle
- β_2 : **eff**et d'**inter**action = différence de pente entre le bras expérimental et le bras contrôle
- γ_k : **eff**et du **bras** sur le risque de sortie d'étude pour chaque cause
- α_k : degré d'**assoc**iation entre QdV et risque de sortie d'étude pour chaque cause

Estimation des modèles

Estimation du modèle conjoint

- ▶ Utilisation du logiciel R, package JM, fonction `jointModel`^a
- ▶ Formes non-paramétriques pour les risques instantanés de base h_0 (M1) et h_{0k} (M2) pour $k = 1, 2$ estimées par des B-splines
- ▶ Méthode adaptative de Gauss-Hermite pour approximer les intégrales sur les effets aléatoires (9 quadratures)
- ▶ Estimation par maximum de vraisemblance de la distribution conjointe du score de QdV et du temps de sorties d'étude



Tous les paramètres des deux sous-modèles sont estimés conjointement

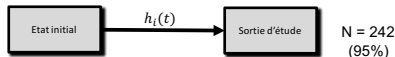
a. D. Rizopoulos, Joint Models for Longitudinal and Time-to-Event Data : With Applications in R. CRC Press, 2012.

Essai clinique PRODIGE5/ACCORD17

Application des modèles conjoints

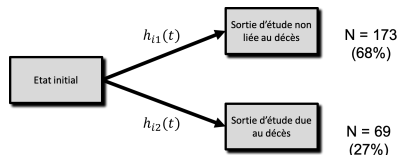
- Analyse de l'échelle **fonction physique**
- Données longitudinales QdV mesurées à 8 visites entre 0 et 36 mois

Modèle 1 sans distinguer les causes



	Pas du tout	Un peu	Assez	Beaucoup
1. Avez-vous des difficultés à faire certains efforts physiques pénibles comme porter un sac à provisions chargé ou une valise ?	1	2	3	4
2. Avez-vous des difficultés à faire une longue promenade ?	1	2	3	4
3. Avez-vous des difficultés à faire un petit tour dehors ?	1	2	3	4
4. Êtes-vous obligé(e) de rester au lit ou dans un fauteuil pendant la journée ?	1	2	3	4
5. Avez-vous besoin d'aide pour manger, vous habiller, faire votre toilette ou aller aux toilettes ?	1	2	3	4

Modèle 2 en distinguant 2 causes possible de sorties d'étude



Résultats

	Modèle 1		Modèle 2	
	Estimation [IC 95 %]	p	Estimation [IC 95 %]	p
Score QdV				
β_1	-0.516 [-0.93 ; -0.10]	0.016	-0.523 [-0.96 ; -0.08]	0.020
β_2	0.069 [-0.42 ; 0.56]	0.779	0.077 [-0.42 ; 0.57]	0.759
Risque de sortie d'étude				
γ	-0.053 [-0.31 ; 0.20]	0.685	-	-
γ_1	-	-	-0.088 [-0.39 ; 0.21]	0.567
γ_2	-	-	-0.110 [-0.46 ; 0.68]	0.704
α	-0.02 [-0.03 ; -0.01]	<0.001	-	-
α_1	-	-	-0.014 [-0.03 ; 0.00]	0.030
α_2	-	-	-0.008 [-0.03 ; 0.02]	0.474

Interprétation des résultats

- Diminution de la QdV au cours du temps ($\hat{\beta}_1$) et pas de différence de QdV entre les deux bras de traitement ($\hat{\beta}_2$)
- Pas de différence entre les deux bras sur le(s) risque(s) de sorties d'étude ($\hat{\gamma}, \hat{\gamma}_1, \hat{\gamma}_2$)

Résultats

	Modèle 1		Modèle 2	
	Estimation [IC 95 %]	p	Estimation [IC 95 %]	p
Score QdV				
β_1	-0.516 [-0.93 ; -0.10]	0.016	-0.523 [-0.96 ; -0.08]	0.020
β_2	0.069 [-0.42 ; 0.56]	0.779	0.077 [-0.42 ; 0.57]	0.759
Risque de sortie d'étude				
γ	-0.053 [-0.31 ; 0.20]	0.685	-	-
γ_1	-	-	-0.088 [-0.39 ; 0.21]	0.567
γ_2	-	-	-0.110 [-0.46 ; 0.68]	0.704
α	-0.02 [-0.03 ; -0.01]	<0.001	-	-
α_1	-	-	-0.014 [-0.03 ; 0.00]	0.030
α_2	-	-	-0.008 [-0.03 ; 0.02]	0.474

Interprétation des résultats

- Diminution de la QdV au cours du temps ($\hat{\beta}_1$) et pas de différence de QdV entre les deux bras de traitement ($\hat{\beta}_2$)
- Pas de différence entre les deux bras sur le(s) risque(s) de sorties d'étude ($\hat{\gamma}, \hat{\gamma}_1, \hat{\gamma}_2$)
- ↘ de 10 points du score de QdV courant correspond :
 - ↗ de $e^{-10\hat{\alpha}} = 22\%$ du risque de sortie d'étude (M1)

Résultats

	Modèle 1		Modèle 2	
	Estimation [IC 95 %]	p	Estimation [IC 95 %]	p
Score QdV				
β_1	-0.516 [-0.93 ; -0.10]	0.016	-0.523 [-0.96 ; -0.08]	0.020
β_2	0.069 [-0.42 ; 0.56]	0.779	0.077 [-0.42 ; 0.57]	0.759
Risque de sortie d'étude				
γ	-0.053 [-0.31 ; 0.20]	0.685	-	-
γ_1	-	-	-0.088 [-0.39 ; 0.21]	0.567
γ_2	-	-	-0.110 [-0.46 ; 0.68]	0.704
α	-0.02 [-0.03 ; -0.01]	<0.001	-	-
α_1	-	-	-0.014 [-0.03 ; 0.00]	0.030
α_2	-	-	-0.008 [-0.03 ; 0.02]	0.474

Interprétation des résultats

- Diminution de la QdV au cours du temps ($\hat{\beta}_1$) et pas de différence de QdV entre les deux bras de traitement ($\hat{\beta}_2$)
- Pas de différence entre les deux bras sur le(s) risque(s) de sorties d'étude ($\hat{\gamma}, \hat{\gamma}_1, \hat{\gamma}_2$)
- ↘ de 10 points du score de QdV courant correspond :
 - ↗ de $e^{-10\hat{\alpha}} = 22\%$ du risque de sortie d'étude (M1)
 - ↗ de $e^{-10\hat{\alpha}_1} = 15\%$ du risque de sortie d'étude non lié au décès (M2)

Etude de simulations

$M = 1000$ jeux de données générés avec $n = 500$ patients chacun

Simulation d'un modèle conjoint avec risques concurrents

- ▶ y_i : scores QdV du patient i , $i = 1, \dots, n$ générés selon un modèle linéaire mixte
- ▶ $T_i = \min(T_{i1}^*, T_{i2}^*, C_i)$, le temps observé pour le patient i avec T_{ik}^* le temps d'événement pour la cause $k = 1, 2$ et C_i le temps de censure

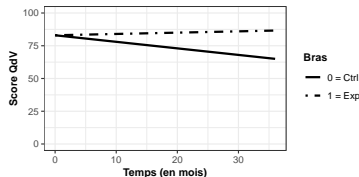
Modèle du temps jusqu'à sortie d'étude pour générer T_{ik}^* :

$$h_{ik}(t) = h_{0k}(t) \exp\{\gamma_k \text{bras}_i + \alpha_k y_i^*(t)\}, \quad k = 1, 2$$

avec $h_{0k}(t) = \phi \lambda t^{\phi-1}$ pour $k = 1, 2$ les risques de base selon une distribution de Weibull, de paramètres ϕ (forme) et λ (échelle) fixés

Description des scénarios (1/2)

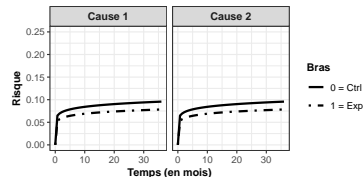
Modèle longitudinal



Qualité de vie qui se détériore pour le bras contrôle contrairement au bras expérimental :

$$\beta_1 = -0.5 \quad \beta_2 = 0.6$$

Modèle temps jusqu'à événement

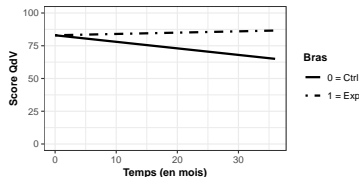


Augmentation du risque similaire selon chaque cause et effet protecteur du bras expérimental :

$$\gamma_1 = \gamma_2 = -0.2$$

Description des scénarios (1/2)

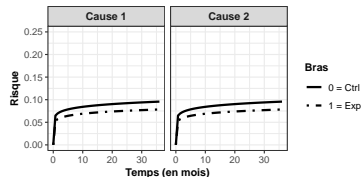
Modèle longitudinal



Qualité de vie qui se détériore pour le bras contrôle contrairement au bras expérimental :

$$\beta_1 = -0.5 \quad \beta_2 = 0.6$$

Modèle temps jusqu'à événement



Augmentation du risque similaire selon chaque cause et effet protecteur du bras expérimental :

$$\gamma_1 = \gamma_2 = -0.2$$

Idées des scénarios

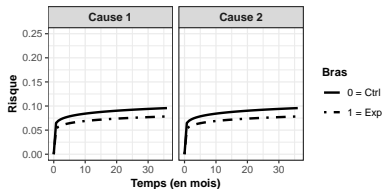
Augmenter le degré d'association entre QdV et risque de sortie d'étude :

⇒ 3 scénarios différents

Description des scénarios (2/2)

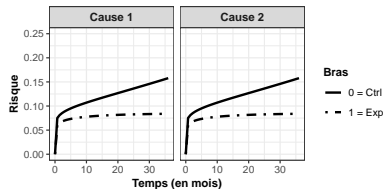
► Scénario 1

$$\alpha_1 = \alpha_2 = 0$$



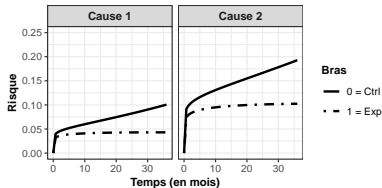
► Scénario 2

$$\alpha_1 = \alpha_2 = -0.02$$



► Scénario 3

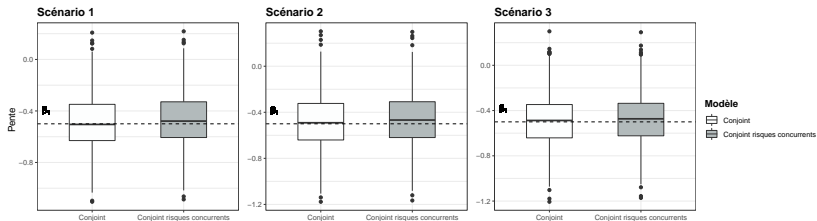
$$\alpha_1 = -0.03, \alpha_2 = -0.02$$



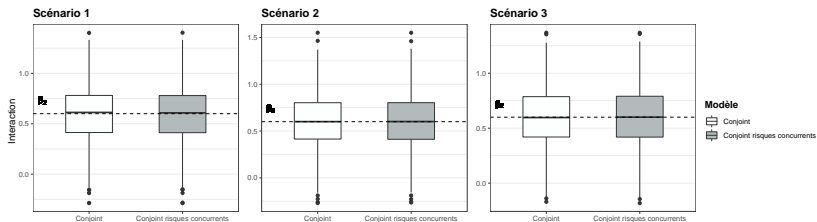
⇒ Terme d'association qui différencie les risques

Résultats sur les paramètres du modèle longitudinal

► Coefficient de pente β_1 :

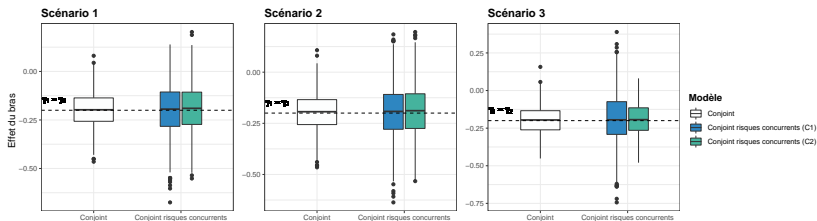


► Coefficient d'interaction β_2 :



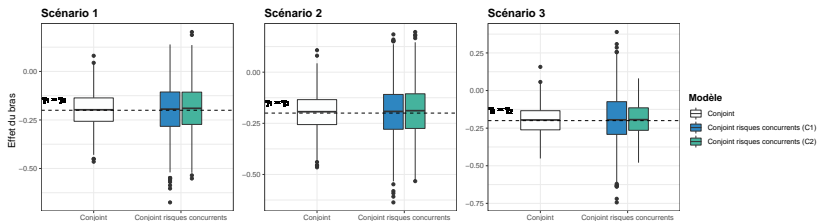
Résultats sur les paramètres du modèle de survie

► Effet du bras sur le risque de sortie d'étude γ , γ_1 et γ_2 :

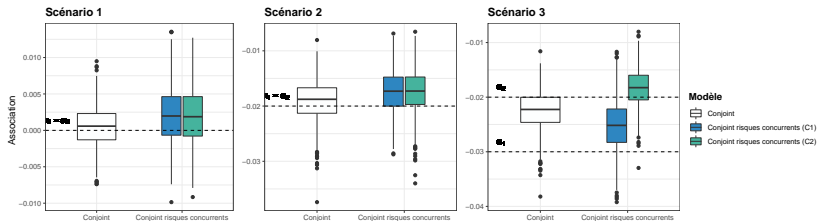


Résultats sur les paramètres du modèle de survie

► Effet du bras sur le risque de sortie d'étude γ , γ_1 et γ_2 :



► Coefficients d'association α , α_1 et α_2 :



Conclusion sur les simulations

Comparaison des 2 modèles

- ▶ Paramètres de pente et d'interaction (β_1 et β_2) + effet du bras sur le(s) risque(s) de sorties d'étude (γ , γ_1 et γ_2) bien estimés par les 2 modèles
- ▶ **Modèle 2** distingue lorsqu'il y a 2 associations différentes selon la cause de sortie d'étude mais semblent mal estimer α_1 et α_2

Conclusion sur les simulations

Comparaison des 2 modèles

- ▶ Paramètres de pente et d'interaction (β_1 et β_2) + effet du bras sur le(s) risque(s) de sorties d'étude (γ , γ_1 et γ_2) bien estimés par les 2 modèles
- ▶ **Modèle 2** distingue lorsqu'il y a 2 associations différentes selon la cause de sortie d'étude mais semblent mal estimer α_1 et α_2

Perspectives

- ▶ Risques de base simulés de Weibull mais estimés avec B-splines :
 ↪ **merlin** (Stata) permet d'estimer selon Weibull
- ▶ **Sensibilité** des résultats et problème de robustesse :
 - Nombre et placement des noeuds pour les B-splines
 - Nombre de quadrature pour approximer les intégrales des effets aléatoires
- ▶ Considérer d'autres **scénarios complémentaires**
 - D'autres paramètres pour les Weibull
 - Ajouter un effet du bras sur le risque de sortie d'étude

Conclusions générales

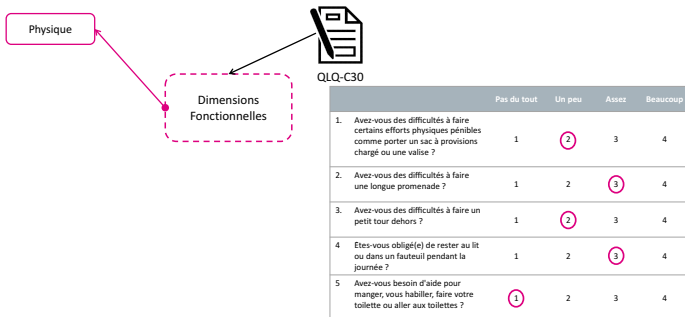
En amont

- ▶ Éviter les données manquantes
- ▶ Lorsqu'elles surviennent → essayer de collecter la raison de sortie d'étude
↪ Considérer différentes causes de sortie d'étude

Lors de l'analyse

- ▶ Modèles conjoints pour QdV en présence de sorties d'étude
- ▶ Modèles à risques concurrents :
↪ Théoriquement mieux mais plus difficile à mettre en oeuvre

Questionnaire EORTC QLQ-C30



► Exemple de la dimension *Fonction Physique* :

- **Variable qualitative** (Pas du tout / Un peu / Assez / Beaucoup)
- **Variable d'intérêt quantitative** (score de 0 à 100)

$$S = \left[1 - \frac{(\frac{1}{J} \sum y_j) - 1}{M - 1} \right] \times 100 = 60$$

Critère d'évaluation lors des simulations

Pour les 3 approches, pour chaque jeu de données m et pour chaque paramètre θ ,

$$\text{Biais}(\theta) = \frac{\hat{\theta}_m - \theta^*}{\theta^*} \times 100$$

- ▶ θ^* la vraie valeur du paramètre θ
- ▶ $\hat{\theta}_m$ l'estimation de θ dans le jeu de données m