

MICHÈLE PICARD FLIBOTTE

**EFFICACITÉ ET INNOCUITÉ D'UN
MÉDICAMENT CONTRE LA MPOC
COMPARATIVEMENT À UN CONTRÔLE**

Essai-stage présenté
à la Faculté des études supérieures de l'Université Laval
dans le cadre du programme de maîtrise en statistique
pour l'obtention du grade de Maître ès sciences (M.Sc.)

DÉPARTEMENT DE MATHÉMATIQUES ET DE STATISTIQUE
FACULTÉ DES SCIENCES ET DE GÉNIE
UNIVERSITÉ LAVAL
QUÉBEC

2011

Table des matières

Table des matières	iii
Liste des tableaux	iv
Table des figures	v
1 Introduction	1
1.1 Contexte	1
1.2 Description de l'étude	2
2 Objectifs	3
2.1 Objectif princial	3
2.2 Objectifs secondaires	3
2.3 Hypothèse	4
3 Méthodes et concepts statistiques	5
3.1 Étude de non-infériorité	5
3.2 Analyse de survie	8
3.2.1 Fonction de survie	8
3.3 Modèle semi-paramétrique de Cox	9
3.3.1 Sélection des variables	11
3.3.2 Variables confondantes	11
3.3.3 Interaction	12
3.3.4 Forme fonctionnelle	12
3.3.5 Hypothèse de proportionnalité	13
3.3.6 Valeurs aberrantes	13
3.3.7 Adéquation du modèle	14
4 Analyse de données	16
4.1 Base de données	16
4.1.1 Analyse exploratoire	17
4.2 Ajustement du modèle de Cox	18
4.2.1 Analyse exploratoire	18

4.2.2	Sélection du modèle	19
4.2.3	Forme fonctionnelle	20
4.2.4	Hypothèse de proportionnalité	21
4.2.5	Diagnostics d'influence (Valeurs aberrantes et données influentes)	22
4.2.6	Adéquation du modèle	23
4.3	Analyse principale	28
4.4	Analyses secondaires	29
4.4.1	Interaction entre les groupes de traitement et chaque covariable	29
4.4.2	Fonction de survie	30
5	Discussion et conclusion	31
6	Annexes	33
6.1	Programmes SAS	33
	Bibliographie	39

Liste des tableaux

4.1	Covariables	17
4.2	Statistiques descriptives pour les covariables	17
4.3	Statistiques descriptives pour la variable d'intérêt et la variable de censure	18
4.4	Analyse univariée des covariables	19
4.5	Résidus de score standardisés, DFBETAS, pour AGE	24
4.6	Résidus de score standardisés, DFBETAS, pour FEV1B	24
4.7	Résidus de déviance	25
4.8	Méthode de sélection des meilleurs sous-ensembles (C_p de Mallows) . .	26
4.9	Ajustement du modèle complet	27
4.10	Ajustement du modèle réduit	27
4.11	Résultat de l'analyse principale	28
4.12	Interaction entre TRT1PN et SEXN	29
4.13	Interaction entre TRT1PN et AGE	29

Table des figures

3.1	Résultats possibles lors d'une étude de non-infériorité	8
4.1	Forme fonctionnelle : Procédure LOESS	20
4.2	Forme fonctionnelle : Énoncé ASSESS dans la procédure PHREG. Valeurs p : AGE=0.3480 et FEV1B=0.5990	20
4.3	Hypothèse de proportionnalité : $\log(-\log(\hat{S}))$ en fonction du temps . .	22
4.4	Hypothèse de proportionnalité : option PH dans ASSESS. Valeurs p : SEYN=0.3430, AGE=0.4720, FEV1B=0.2000 et TRT1PN=0.9460	23
4.5	LD et LMAX	25
4.6	Résidus du score	25
4.7	C_p de Mallows en fonction du nombre de variables	26
4.8	Fonction de survie	30

Chapitre 1

Introduction

1.1 Contexte

Quintiles Canada Inc., la compagnie chez qui j'ai fait mon stage, reçoit des contrats pour réaliser des études cliniques afin de tester de nouveaux médicaments ou de nouvelles molécules. Ces contrats sont habituellement accordés par des compagnies pharmaceutiques qui font de la recherche afin de commercialiser de nouveaux médicaments. Avant qu'un nouveau médicament ne sorte sur le marché, il doit être approuvé par l'organisme de réglementation du pays en question. Au Canada, c'est Santé Canada qui s'assure que les nouveaux médicaments sont suffisamment sécuritaires et efficaces pour être mis en marché. Aux États-Unis, c'est la FDA (U.S. Food and Drug Administration) qui exerce ce rôle. Comme la majorité des compagnies pharmaceutiques partenaires sont américaines, elles sont régies par la FDA et par conséquent, Quintiles Canada Inc. l'est aussi. Certaines décisions prises lors d'un projet sont du fait expliquées par ceci. Comme il y a beaucoup de temps et d'argent en jeu dans le processus de recherche, le respect de la confidentialité est crucial. Pour cette raison, tout au long de ce rapport de stage, aucun nom de compagnie pharmaceutique ni de médicament ne seront mentionnés afin de les garder confidentiels.

Dans le cadre de ce stage, le médicament à l'étude pourrait servir de traitement pour les symptômes de la maladie pulmonaire obstructive chronique (MPOC), une maladie qui englobe entre autres la bronchite chronique et l'emphysème. C'est une maladie de plus en plus répandue dans la population. En 2005, plus de 750 000 Canadiens et Canadiennes étaient touchés par la MPOC. Elle est habituellement causée par le tabagisme et a pour symptômes la toux, la sécrétion de mucus et l'essoufflement. C'est une maladie qui ne se guérit pas, mais dont les symptômes se traitent. Afin de traiter l'essouffle-

ment, les médecins prescrivent des bronchodilatateurs soit de type beta-2-agoniste soit de type anticholinergique (2).

1.2 Description de l'étude

Étude clinique multi-centre de phase III, randomisée, à double insu, avec groupe contrôle ayant pour but d'évaluer l'efficacité et l'innocuité de l'utilisation à long terme d'un médicament dans le traitement de patients atteints de la MPOC de niveau modéré à sévère. Les sujets à l'étude, hommes ou femmes, doivent être âgés d'au moins 40 ans, doivent être diagnostiqués pour la MPOC de niveau modéré à sévère sans l'être pour l'asthme et doivent avoir eu au moins une exacerbation de leur maladie au cours de la dernière année. Ils doivent également avoir des antécédents de fumeurs, c'est-à-dire avoir fumé au moins 15 paquets-année.

À l'étape de la randomisation (2^e visite), les patients recrutés pour participer à cette étude clinique se voient aléatoirement attribués un des deux traitements : le traitement à l'étude ou un contrôle. Tous deux sont administrés par pulvérisation. Le contrôle est un médicament déjà reconnu comme traitement standard de la MPOC. L'étude est réalisée à double insu, ce qui signifie que ni le patient ni l'équipe médicale ne connaissent le traitement utilisé. Jusqu'à ce que l'étude soit terminée, les analyses intérimaires sont à l'aveugle, c'est-à-dire que l'équipe de statisticiens qui travaille directement ou indirectement avec l'équipe médicale ne connaissent pas le vrai traitement utilisé. Donc les analyses qui ont été réalisées, entre autres dans le cadre de ce stage, donnent des résultats partiellement bidons étant donnés les traitements bidons. Toutefois, la démarche pourra être utilisée à la fin de l'étude. Les patients qui participent à l'étude sont suivis sur une période d'un an grâce à six visites espacées de trois mois. La première visite sert à déterminer si le patient est admissible et à recueillir les données de base qui seront utilisées à des fins de comparaison avec les données recueillies aux visites subséquentes. Il a été calculé que pour détecter une différence significative entre le traitement et le contrôle, 86 événements principaux devront être observés. Seulement une fois ce nombre atteint, l'étude se terminera. L'évènement principal est défini comme étant une hospitalisation due à une aggravation des symptômes de la MPOC ou la mort respiratoire, la première à survenir. Le nombre nécessaire n'avait pas encore été observé au moment de la rédaction de cet essai.

Chapitre 2

Objectifs

2.1 Objectif principal

L'objectif principal de cette étude est de déterminer l'incidence et le risque relatif de l'évènement principal associés à l'utilisation du médicament comparativement à celle du contrôle. Afin d'évaluer cet objectif, on étudiera les relations existantes entre le temps entre la randomisation et l'évènement principal et certaines variables recueillies lors de l'étude clinique.

2.2 Objectifs secondaires

Bien qu'une étude soit principalement réalisée pour évaluer un objectif principal, les objectifs secondaires sont souvent d'un très grand intérêt pour les chercheurs qui commandent l'étude. Ici, les objectifs secondaires sont de déterminer l'influence du médicament sur l'exacerbation de la MPOC, sur les effets secondaires et sur la mortalité de même que d'évaluer les changements dans les symptômes et le fonctionnement des patients atteints de la MPOC et qui prennent le médicament. Dans le cadre de ce stage, seulement le premier objectif sera évalué. Les critères d'évaluation sont divisés en deux types : l'incidence d'un évènement et le temps entre la randomisation et ce même évènement. On étudiera l'évènement principal (seulement pour l'incidence car le temps entre la randomisation et l'évènement principal fait l'objet de l'objectif principal), les évènements d'aggravation de la MPOC et la mort de toutes causes. Le deuxième objectif serait évalué par le changement entre la visite de base et les autres visites pour

la capacité inspiratoire et pour le résultat à un questionnaire portant sur le statut de santé.

2.3 Hypothèse

L'hypothèse de l'étude est une hypothèse de non-infériorité qui signifie que, comparativement au contrôle, l'administration du médicament n'entraînera pas d'augmentation de l'incidence d'évènements principaux.

Chapitre 3

Méthodes et concepts statistiques

3.1 Étude de non-infériorité

Il existe quatre grands types de groupes contrôle dans une étude clinique : aucun traitement, placebo, relation dose-effet et substance active connue. Dans les deux premiers cas, il est approprié de tester si le traitement à l'étude est supérieur au contrôle ; c'est ce qu'on appelle une étude de supériorité. Par définition, un placebo est un traitement « factice » qui paraît aussi identique que possible au traitement étudié quant à ses caractéristiques physiques comme la couleur, le poids, le goût et l'odeur, mais qui ne contient pas le médicament étudié (9). Selon ces mêmes lignes directrices, lorsqu'un traitement efficace connu existe pour traiter une maladie dans une population particulière, il n'est éthiquement pas recommandé de faire un essai contrôlé contre un placebo. Dans le troisième cas, aucun test n'est habituellement effectué. Il est plutôt utilisé dans une étude clinique de phase I pour déterminer la dose maximale tolérée qui sera utilisée lors d'une étude subséquente de phase II. Dans le dernier cas, si on ne s'intéresse qu'à tester si le nouveau traitement est supérieur au contrôle et que l'on a de bonnes raisons de le croire, on peut faire une étude de supériorité. Dans toute autre situation, c'est-à-dire si le but de l'étude est de vérifier si la différence entre le traitement et le contrôle est suffisamment petite pour affirmer que le traitement n'est pas inférieur au contrôle ou si on n'a aucune idée si le traitement est supérieur ou inférieur au contrôle, il est préférable de faire une étude de non-infériorité (13). C'est au moment de la planification qu'il faut prendre cette décision et à ce même moment il est possible de formuler une hypothèse de supériorité dans le cas où le traitement à l'étude s'avérait non-inférieur au contrôle. Bien que ce type d'étude soit un peu plus complexe au niveau de l'interprétation, il a l'avantage qu'en plus de tirer des conclusions typiques à ce type d'études, on peut arriver aux mêmes conclusions que l'on aurait obtenues dans le cadre

d'une étude de supériorité, si non-infériorité il y a. On peut également effectuer des études d'équivalences dans une situation où on veut faire approuver un médicament générique. Dans un tel cas, il faut montrer que le médicament générique contient les mêmes ingrédients médicamenteux que le médicament original et ce, en même quantité et de même qualité. Il faut également prouver que le rythme d'absorption est le même (8).

On voit parfois des études de non-infériorité réalisées dans des situations où un médicament expérimental aurait certains bienfaits (meilleure toxicité, mieux toléré, moins d'effets secondaires, etc.) comparativement à un autre traitement et on veut s'assurer de ne pas trop perdre d'efficacité. Car du point de vue médical, il peut être acceptable que le médicament à l'étude soit un peu moins efficace s'il a d'autres bienfaits.

Rappelons la formulation des hypothèses pour un test unilatéral de supériorité afin de mieux comprendre la différence par rapport au test de non-infériorité.

$$\begin{array}{ccc} H_0 : T \leq C & & H_0 : T - C \leq 0 \\ & \iff & \\ H_1 : T > C & & H_1 : T - C > 0 \end{array}$$

où T est le traitement à l'étude et C est le contrôle (placebo)

On rappelle également que les hypothèses sont formulées de façon à arriver à la conclusion voulue en refusant H_0 . Comme on veut arriver à la conclusion que T est supérieur à C, on doit mettre l'inéquation $T > C$ dans la contre-hypothèse. Dans le cadre de cette analyse, on s'intéresse au risque relatif du nombre d'événements avec le médicament à l'étude par rapport au contrôle. Le test de supériorité serait alors ainsi formulé :

$$\begin{array}{l} H_0 : HR\left(\frac{T}{C}\right) \geq 1 \\ H_1 : HR\left(\frac{T}{C}\right) < 1 \end{array}$$

Dans le cadre d'une étude de non-infériorité, au lieu d'exiger que $T > C$, on est prêt à accepter une certaine marge d'infériorité entre le médicament et le contrôle tel que $T > C - M$. Un peu de manipulations algébriques nous permet de réécrire cette inéquation :

$$T > C - M \iff T - C > -M \iff C - T < M$$

Lorsque cette inéquation est vérifiée, ça signifie que T est non-inférieur à C et c'est la conclusion à laquelle on veut arriver. Il faut donc formuler le test d'hypothèse de façon à refuser H_0 pour conclure que T est non-inférieur.

$$\begin{array}{l} H_0 : C - T \geq M \\ H_1 : C - T < M \end{array}$$

En terme de risque relatif, on obtient les hypothèses suivantes :

$$H_0 : HR\left(\frac{C}{T}\right) \leq 1 + M$$

$$H_1 : HR\left(\frac{C}{T}\right) > 1 + M$$

Par contre, comme nous sommes intéressés au risque relatif de T par rapport à C, il faut transformer nos hypothèses pour obtenir le test d'hypothèse final suivant :

$$H_0 : HR\left(\frac{T}{C}\right) \geq 1 + M$$

$$H_1 : HR\left(\frac{T}{C}\right) < 1 + M \text{ (Si on rejette } H_0 : T \text{ est non-inférieur au contrôle)}$$

Il s'en suit une décision importante à prendre lors de la planification d'une étude de non-infériorité. Il faut choisir la marge d'infériorité acceptable, M, pour conclure que le médicament à l'étude est encore suffisamment efficace. Dans une étude de supériorité, M=0. Dans le cadre de cette étude clinique, basé sur les connaissances cliniques du médicament, la marge d'infériorité a été fixée à 40%. Comme il est mentionné précédemment, au moment de la planification de l'étude, il faut formuler le test d'hypothèse de non-infériorité et également celui de supériorité dans le cas où T s'avérait non-inférieur à C. Ce qui donne les tests suivants :

1. Test de non-infériorité	2. Test de supériorité
$H_0 : HR\left(\frac{T}{C}\right) \geq 1.4$	$H_0 : HR\left(\frac{T}{C}\right) \geq 1$
$H_1 : HR\left(\frac{T}{C}\right) < 1.4$	$H_1 : HR\left(\frac{T}{C}\right) < 1$

Afin d'effectuer le test d'hypothèse, on doit comparer la borne supérieure de l'intervalle de confiance (IC) aux bornes qui nous intéressent. La figure 3.1 montre les résultats possibles lors de cette étude et voici l'interprétation que l'on devrait en faire :

1. La borne supérieure de l'IC est < 1.4 , donc il y a non-infériorité. Bien que l'estimé ponctuel suggère l'égalité des traitements, on ne peut pas conclure à l'égalité.
2. La borne supérieure de l'IC est > 1.4 , donc la non-infériorité n'est pas démontrée. L'estimé favorise C.
3. La borne supérieure de l'IC est au-dessus de 1.4, donc la non-infériorité n'est pas démontrée. L'estimé est 1 ce qui suggère l'égalité des traitements.
4. La borne supérieure de l'IC est < 1.4 , donc il y a non-infériorité. La supériorité n'est pas démontrée car la borne supérieure de l'IC est > 1 . L'estimé favorise T.
5. En plus d'y avoir non-infériorité, la supériorité de T est démontrée car la borne supérieure de l'IC est < 1 .

6. La borne supérieure de l'IC est < 1.4 , donc il y a non-infériorité de T. Par contre, l'estimé favorise C et C est significativement supérieur à T car la borne inférieure de l'IC est > 1 . Ce qui contre-dit le premier résultat. Ne devrait pas se produire.

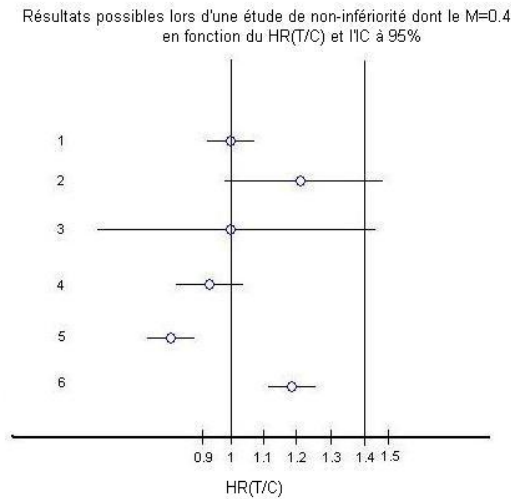


FIG. 3.1 – Résultats possibles lors d'une étude de non-infériorité

3.2 Analyse de survie

3.2.1 Fonction de survie

Soit X une variable aléatoire non-négative provenant d'une population homogène. X représente le temps jusqu'à ce qu'un événement précis se produise, par exemple une récidence à un cancer, le décès... La fonction de survie est la probabilité qu'un individu n'ait pas observé cet événement précis, ou survive, au temps x . On emploie souvent le terme "survivre" même si l'événement n'est pas nécessairement le décès. Elle est définie comme

$$S(x) = Pr(X > x), \quad x > 0.$$

Elle est estimée par l'estimateur de Kaplan-Meier qui est défini comme suit pour toutes les valeurs t pour lesquelles il y a des données :

$$\hat{S}(x) = \begin{cases} 1 & \text{si } x < x_1 \\ \prod_{x_i \leq x} [1 - \frac{d_i}{Y_i}] & \text{si } x_1 \leq x, \end{cases}$$

où x_i est le temps auquel il s'est produit d_i événements et Y_i est le nombre d'individus à risque au temps x_i .

Dans SAS, la procédure utilisée pour estimer la fonction de survie par la méthode de Kaplan-Meier est `lifetest`. L'énoncé `strata` permet d'estimer la fonction de survie séparément pour chaque strate. Il permet également de tester s'il y a une différence entre les fonctions de survie de chacune des strates grâce à différents tests. Le test utilisé pour les analyses est celui du log-rang selon $H_0 : S_1(x) = S_2(x)$ pour tout temps x où un événement s'est produit. Ce qui revient à tester que les groupes sont homogènes :

$$\frac{\sum_{i=1}^D d_{i1} - \frac{Y_{i1}}{Y_{i1}+Y_{i2}}}{\sqrt{\sum_{i=1}^D \frac{Y_{i1}Y_{i2}}{(Y_{i1}+Y_{i2})^2}}} \approx N(0, 1),$$

où $S_1(x)$ représente la survie dans le groupe 1 et $S_2(x)$ celle dans le groupe 2, Y_{i1} est le nombre d'individus à risque au temps x_i dans le groupe 1 et Y_{i2} est celui au temps x_i dans le groupe 2 et d_{i1} est le nombre d'événements observés au temps x_i dans le groupe 1.

3.3 Modèle semi-paramétrique de Cox

Lorsqu'on parle de modélisation que ce soit d'un modèle de régression linéaire ou semi-paramétrique de Cox, il existe deux lignes de pensées quant au moment de choisir les variables à inclure dans le modèle. Certains prônent la sélection *a priori* et d'autres, la sélection *a posteriori* par rapport aux observations recueillies.

L'hypothèse nécessaire lors d'une sélection *a priori* est que la liste exhaustive des variables qui expliquent l'événement soit connue dès le départ. Ceci doit se faire avant même le début de l'étude clinique et de la collecte des données. Rendu aux analyses c'est cette liste qui sera incluse dans le modèle. Or, dans la majorité des cas, cette liste n'est pas nécessairement connue. Elle peut être supposée et dans ce cas il pourrait y avoir un risque de biais d'omission de certains paramètres.

Ceux qui prônent la sélection *a posteriori* partent avec l'idée qu'ils ne connaissent pas cette liste exhaustive. Ils incluent donc toutes les variables qui expliquent potentiellement la variable étudiée et utilisent une méthode de sélection quelconque pour déterminer quelles variables l'expliquent vraiment. Un des inconvénients de cette méthode est que la sélection des variables est basée sur les données observées lors de cette étude clinique précise et qu'ensuite, on utilise ces mêmes données pour faire l'analyse. La méthode idéale serait de faire une première étude afin de déterminer les variables explicatives significatives et ensuite en réaliser une deuxième qui utiliserait ce modèle pour en faire l'analyse. Dans la réalité, il est impensable de réaliser deux études pour

obtenir un seul résultat. C'est pourquoi la sélection est réalisée sur les mêmes données que l'analyse. C'est également pourquoi certains prônent la méthode *a priori*. Comparativement à la méthode précédente, la méthode *a posteriori* est plus biaisée car en plus du biais d'omission, il y a un risque potentiel de biais de sélection. Par contre, la variance est diminuée. Bref, il y a des risques à utiliser l'une ou l'autre de ces méthodes ; il faut seulement en être conscient lors de la prise des décisions et choisir celle qui nous convient le mieux.

Comme mentionné précédemment, Quintiles Canada Inc. doit se conformer aux lignes directrices de la FDA afin de produire ses analyses. Un de leurs principes consiste à décider, dès les premières étapes de la planification d'une étude clinique, les variables qui seront utilisées dans le modèle semi-paramétrique de Cox. C'est donc aux spécialistes en santé de déterminer quelles variables expliquent l'évènement principal avant même de recueillir les données. Ils se basent habituellement sur des études antérieures réalisées sur le même évènement. Est-ce qu'ils les ont toutes incluses ? Qui sait ? Dans le cadre de ce stage, on a décidé de faire une sélection de variables *a posteriori* bien que les analyses publiées à la toute fin de l'étude clinique utiliseront les variables sélectionnées *a priori*.

On définit le modèle semi-paramétrique de Cox comme suit :

$$\lambda(x) = \lambda_0(x)e^{\beta_1 Z_1 + \beta_2 Z_2 + \dots + \beta_p Z_p}$$

où $\lambda_0(x)$ est un risque relatif de base arbitraire et Z_i sont les covariables indépendantes de X .

La procédure SAS utilisée pour ajuster un modèle de Cox aux données est PHREG. Cette procédure utilise la méthode de vraisemblance partielle pour estimer les paramètres du modèle. Elle suppose qu'il n'y a pas d'évènements qui se produisent en même temps. Puisque cette hypothèse est plus ou moins réaliste, plusieurs méthodes existent pour calculer la vraisemblance partielle tout en tenant compte de ces égalités. Il y a entre autres la méthode exacte qui ordonne les évènements égaux de toutes les façons possibles et calcule la vraisemblance partielle pour chacune d'elles. Cette méthode peut s'avérer fastidieuse avec un nombre élevé d'égalités. Une des méthodes recommandées dans la littérature est la méthode d'Efron. Elle est une approximation de la méthode exacte dont les résultats sont assez près, mais dont le temps d'exécution est beaucoup plus raisonnable lorsqu'on travaille avec de larges bases de données. C'est cette méthode qui a été préconisée.

3.3.1 Sélection des variables

Pour sélectionner les variables qui expliquent l'événement principal, il existe plusieurs méthodes de sélection *a posteriori* automatisées dans SAS telles que l'exclusion, l'inclusion, la méthode pas-à-pas, la sélection par le score. Chacune d'elles procède de façon différente pour arriver avec un modèle construit par les variables significatives au seuil exigé. Par exemple, la méthode de l'exclusion part avec le modèle contenant toutes les variables disponibles. Ensuite, une à une il enlève les variables qui ne sont pas significatives au seuil exigé. C'est la seule condition nécessaire pour qu'une variable soit exclue du modèle.

Nous exigerons un seuil observé de 5%. Un des inconvénients de ces méthodes est qu'aucune d'elles ne prend en considération la confondance potentielle entre les variables. Car en plus d'exiger qu'une variable soit significative au seuil de 5%, elle doit également ne pas être confondante avec le traitement pour entrer dans le modèle. On veut s'assurer que les covariables ne sont pas confondantes avec le traitement car l'objectif principal de cette étude est de déterminer le risque relatif de l'évènement principal associée au traitement. Donc on s'intéressera seulement à la confondance des covariables par rapport au traitement. Par conséquent, la sélection des variables sera faite de façon manuelle afin d'inclure les variables confondantes. La section 3.3.2 explique comment déterminer qu'une variable est confondante.

On procédera à la sélection des variables à partir du modèle contenant toutes les variables disponibles et on se basera sur le niveau de significativité obtenu lors de l'analyse univariée pour déterminer l'ordre des variables à conserver ou à exclure. La variable la moins significative sera vérifiée en premier. Si elle n'est pas confondante avec la variable de traitement et qu'elle n'est pas significative, elle sera éliminée du modèle ; sinon elle sera conservée.

3.3.2 Variables confondantes

Pour déterminer qu'une variable est confondante, par exemple **SEXN**, on regarde le pourcentage de changement que provoque le retrait de **SEXN** par rapport à l'estimation du paramètre de la variable de traitement. On calcule le pourcentage de changement de la façon suivante :

$$\Delta = \frac{100(\hat{\beta}_{1\text{réduit}} - \hat{\beta}_{1\text{complet}})}{\hat{\beta}_{1\text{complet}}},$$

où β_1 = paramètre de TRT1PN. Tout changement supérieur à 15% est considéré important et **SEXN** sera conservée dans le modèle même si elle n'est pas significative au seuil de 5%.

3.3.3 Interaction

Ensuite, il faut s'assurer que s'il y a une interaction entre certaines variables, elle sera incluse dans le modèle. Pour ce faire, on ajoute une à une l'interaction entre deux variables et on vérifie si elle est significative au seuil de 5%.

3.3.4 Forme fonctionnelle

Une fois que le modèle est déterminé, il faut vérifier la forme fonctionnelle des variables continues. Pour une covariable, on veut trouver la forme fonctionnelle qui explique le mieux son effet sur la survie obtenue par le modèle de Cox. Dans le cadre de ce stage, deux méthodes utilisant les résidus de martingale ont été utilisées : la procédure LOESS de SAS et la commande ASSESS dans la procédure PHREG. Les résidus de martingale représentent la différence entre le nombre d'événements observés et le nombre d'événements espérés selon le modèle à chaque temps d'événement observé.

La première méthode est moins directe et plus subjective que la seconde. Il faut commencer par calculer les résidus de martingale du modèle de Cox comprenant toutes les covariables excepté celle étudiée et tracer le graphique des résidus de martingale en fonction de cette variable. Afin de faciliter l'interprétation du graphique, on utilise une courbe de lissage obtenue grâce à une régression spline de la procédure LOESS de SAS. Si on obtient une droite, on pourra conserver la variable sous sa forme continue. Sinon, on devra la transformer. Tous les types de transformations connues peuvent être utilisées telles que quadratique, cubique, catégorique, logarithmique, ...

La seconde méthode produit un graphique des résidus cumulatifs de martingale par rapport à la covariable elle-même. Elle produit ce graphique et la valeur p du test du supremum de type Kolmogorov basé sur une simulation de 1000 répétitions selon H_0 : la forme fonctionnelle est bien spécifiée (5).

3.3.5 Hypothèse de proportionnalité

Une hypothèse importante du modèle de Cox est que le risque relatif soit constant par rapport au temps. Cette fois-ci, on a utilisé trois méthodes pour s'assurer que l'hypothèse est respectée.

La première s'applique seulement aux variables continues. Si l'hypothèse de proportionnalité était violée pour une variable, ça voudrait dire que l'effet de cette variable varie avec le temps. La solution à ce problème serait d'ajouter une nouvelle variable qui serait son interaction avec le temps. Ensuite, on vérifie si cette nouvelle variable est significative. Si elle ne l'est pas, on ne peut pas conclure que son effet varie dans le temps et donc rien ne dit que l'hypothèse de proportionnalité n'est pas respectée.

La seconde méthode concerne les variables discrètes seulement. On trace le graphique du $\log(-\log(\hat{S}))$ en fonction du temps où \hat{S} est la probabilité de survie estimée pour chaque sous-groupe. Un sous-groupe correspond à une valeur donnée de la covariable. Par exemple, pour la variable de sexe, le graphique comprendra deux courbes, une pour les hommes et une pour les femmes. Pour que l'hypothèse de proportionnalité soit respectée, les courbes doivent être verticalement parallèles.

La dernière méthode permet de vérifier l'hypothèse de proportionnalité pour toutes les covariables, discrètes et continues. Elle trace le graphique du processus du score normalisé (Standardized Score Process) en fonction du temps jusqu'à l'événement et ce, pour chaque covariable. Une valeur p est également calculée par le test supremum du type Kolmogorov basé sur une simulation de 1000 répétitions selon H_0 : l'hypothèse de proportionnalité est respectée (5).

3.3.6 Valeurs aberrantes

Une autre vérification importante à faire lorsqu'on modélise des données concerne les valeurs aberrantes et les données influentes. La procédure PHREG produit certaines statistiques qui peuvent être utilisées à ces fins telles que `DFBETAS`, `LD`, `LMAX`, `RESDEV` et `RESSCO`.

La statistique `DFBETAS`, résidu de score standardisé, mesure le changement dans l'estimation du paramètre de chaque covariable continue ($\hat{\beta}_i - \hat{\beta}_{ij}$) lorsque la j^e observation est éliminée des données. On obtient une valeur par paramètre i et par observation. Belsey, Kuh et Welsch (4) suggèrent d'utiliser la borne suivante pour identifier les données

potentiellement influentes :

$$|DFBETA_{ij}| > 2\sqrt{\frac{1}{n}}$$

LD mesure le changement du log de la vraisemblance partielle lorsqu'une observation est éliminée des données et LMAX mesure l'influence relative des observations. Ces deux statistiques permettent de vérifier l'impact de chaque observation sur l'ajustement général du modèle.

Les résidus de déviance, RESDEV peuvent être utilisés pour détecter les valeurs aberrantes et pour vérifier l'adéquation du modèle. Nous les utiliserons seulement pour détecter les valeurs extrêmes. Ces résidus sont dérivés pour chaque observation. Ils ont une distribution symétrique autour de 0 et un écart-type de 1. On portera une attention particulière aux observations ayant un résidu de déviance supérieur à 3.

Ensuite, RESSCO sont les résidus du score. Ils sont calculés pour chaque covariable continue et chaque observation. Ils permettent également de détecter les valeurs extrêmes. En traçant le graphique de ces résidus en fonction de la covariable on identifie les patients qui ont des résidus élevés.

À l'aide de ces cinq statistiques, on identifie les observations qui sont possiblement aberrantes ou influentes et on s'assure qu'il n'y a pas d'erreur dans la saisie des données. Si les valeurs sont exactes, il est possible que le modèle ne s'ajuste pas bien aux données. Dans un tel cas, il n'est pas recommandé d'exclure ces données car elles peuvent contenir de l'information importante sur l'évènement étudié. Si on décidait de les exclure, il serait très important de bien documenter et expliquer les raisons qui motivent ces exclusions. Il faudrait aussi indiquer les limitations des résultats. Selon la littérature (10), si le pourcentage de données influentes n'excède pas 5%, il n'y a pas de raisons de s'inquiéter.

3.3.7 Adéquation du modèle

Le C_p de Mallows est un indicateur de biais. Il est défini comme suit :

$$C_p = \frac{1}{\hat{\sigma}^2} RSS_p - n + 2p,$$

où $\hat{\sigma}^2$ est une estimée de σ^2 et RSS_p est la somme des carrés résiduels (6). Il est habituellement utilisé afin de sélectionner un modèle non biaisé parmi tous les modèles possibles grâce à l'option `selection=SCORE` de SAS. À cette étape, nous utiliserons cette option pour vérifier que le modèle obtenu est bon et qu'il n'est pas biaisé. SAS

considère toutes les combinaisons possibles de covariables qui pourraient former un modèle avec p covariables et calcule le C_p de Mallows pour chacun d'eux. Au total, il y a 2^p sous-ensembles possibles. Avec l'option `best=3`, les 3 meilleurs sous-ensembles avec p covariables seront imprimés. Mallows suggère que le meilleur modèle est celui où $C_p \leq p$ avec p le plus petit possible (6). Il mentionne que ce n'est pas toujours vrai, en particulier lorsque beaucoup de sous-ensembles ont cette propriété. Dans un tel cas, il faudrait se baser sur les notions de sujet-matière pour identifier le meilleur modèle. Pour notre utilisation, on vérifiera que le modèle obtenu par la méthode manuelle nous donne un $C_p \leq p$ pour s'assurer qu'il n'est pas biaisé.

Ensuite, il faut vérifier l'adéquation du modèle. Dans la procédure `PHREG`, il n'existe aucune option qui permette de le faire directement ; il faut le faire manuellement. Pour comparer le modèle complet et le modèle choisi, on effectuera un test du rapport des vraisemblances dont les hypothèses sont :

H_0 : le modèle réduit est adéquat

H_1 : le modèle complet est adéquat

Pour ce faire, on ajuste chacun des deux modèles séparément, et on calcule la statistique $LRT = |L_1 - L_2|$ où L_x est le $-2\log L$ obtenu lors de l'ajustement de chacun des modèles. Comme on prend la valeur absolue, l'ordre de L_1 et L_2 n'a pas d'importance. Cette statistique suit approximativement une distribution khi carré dont les degrés de liberté représentent la différence entre le nombre de paramètres de chaque modèle. De cette façon, on peut calculer la valeur p associée à ce test.

Si la valeur p n'est pas significative, alors le modèle réduit est adéquat. Si la valeur p est significative, ça pourrait vouloir dire que le modèle réduit simplifie trop le modèle complet et dans ce cas, la modèle complet serait préférable.

Chapitre 4

Analyse de données

Chez Quintiles Canada Inc., différentes équipes travaillent conjointement à mener à terme les projets. Une équipe clinique s'occupe de tout ce qui est en lien avec les patients et les traitements. Ensuite, l'équipe de gestion de données reçoit les données telles qu'elles ont été saisies par l'équipe médicale et les traite de façon à ce que nous, l'équipe de biostatistique, puissions faire les analyses statistiques.

4.1 Base de données

Lorsque les données nous parviennent, elles sont dans des fichiers SAS, un fichier différent par panel. Habituellement, les panels sont les suivants : données démographiques des patients (âge, sexe, origine, ...), historique médical, signes vitaux, effets secondaires,... Afin de produire les analyses du projet qui m'a été assigné, seules les données démographiques et les effets secondaires seront utilisés.

Au total, environ 800 patients ont été recrutés pour participer à l'étude dont 43% de femmes et 57% d'hommes. Un patient est de sexe inconnu et a été exclu des analyses. Plusieurs variables démographiques ont été recueillies, mais il a été décidé d'utiliser seulement l'âge, le sexe, le statut de fumeur, l'indice de masse corporelle (IMC) et le volume expiré en une seconde (FEV_1) à l'enrôlement comme covariables. Comme mentionné plus tôt, ces variables ont été choisies *a priori* par les experts en santé et seront utilisées pour l'analyse finale. Pour les fins de ce stage, on effectuera une sélection manuelle pour inclure seulement celles qui sont significatives ou confondantes avec le traitement. Afin de faire partie des analyses, un patient devait avoir fourni une valeur

4.2 Ajustement du modèle de Cox

En rappel, l’analyse principale de ce projet est de déterminer l’incidence et le risque relatif de l’évènement principal associés à l’utilisation du médicament comparativement à celle du contrôle. Pour ce faire, on ajustera un modèle de Cox pour vérifier l’hypothèse de non-infériorité des deux groupes de traitement. Avant toute chose, on a besoin d’une variable contenant le temps entre la randomisation et l’évènement principal et d’une variable de censure. L’évènement principal est défini comme étant une hospitalisation due à une aggravation des symptômes de la MPOC ou la mort respiratoire, la première à survenir. Or, dans la base de données, ces deux évènements sont identifiés dans des variables distinctes avec des dates indépendantes. Il faut donc les regrouper en une seule. On crée donc `TIM2PRMD` qui est le temps en jours entre la randomisation et l’évènement principal et `CENSPRM` qui est la variable de censure. Si un patient n’a vécu aucun des deux évènements et n’est pas décédé d’une autre cause durant l’étude, `CENSPRM` sera mis à 0 et `TIM2PRMD` sera mis à la date de la fin de l’étude. Par contre, s’il est décédé pour une autre raison, `TIM2PRMD` sera la date du décès et `CENSPRM` sera 0. Finalement, si le patient a eu une aggravation des symptômes de la MPOC ou une mort respiratoire, `CENSPRM` sera 1 et `TIM2PRMD` sera la date du premier évènement à survenir.

4.2.1 Analyse exploratoire

Comme première analyse, on a calculé quelques statistiques descriptives pour les deux nouvelles variables. Elles sont résumées dans le tableau 4.3. Seulement 9.34% des

TAB. 4.3 – Statistiques descriptives pour la variable d’intérêt et la variable de censure

Variable	N	Moyenne	Écart-type	Min	Q1	Médiane	Q3	Max
TIM2PRMD	≈ 800	250.87	115.06	1	155	271	360	494
						CENSPRM		
						0	1	
				Proportion	90.66	9.34		

patients ont eu un évènement. À ce stade, on peut penser avoir certains problèmes lors de l’ajustement du modèle de Cox étant donné le faible nombre d’évènements observés. Dès le début de l’étude on s’attendait à observer une faible proportion d’évènements, c’est pourquoi la taille d’échantillon prévue est très grande. Avec 86 évènements, ce sera probablement suffisant pour détecter une différence entre les traitements. On peut

aussi penser avoir des problèmes avec la puissance de l'étude si le nombre d'événements n'est pas aussi élevé que prévu. Au moment de la planification d'une étude, il faut déterminer la taille d'échantillon nécessaire et l'un des paramètres entrant dans ce calcul est le niveau de puissance désiré. On calcule le nombre d'événements nécessaires pour atteindre cette puissance. Pour cette étude, on a calculé que 86 événements sont nécessaires. Donc l'étude se terminera seulement après avoir observé 86 événements. De cette façon, on s'assure qu'à la fin de l'étude la puissance sera respectée. Il est toutefois possible qu'au moment des analyses réalisées dans le cadre de ce stage, la puissance ne l'était pas.

4.2.2 Sélection du modèle

Avant de commencer la sélection du modèle, on a fait une analyse univariée des 6 covariables d'intérêt. Le tableau 4.4 montre les valeurs p obtenues pour chacune des covariables.

À partir du modèle contenant toutes les variables disponibles, on enlève la variable

TAB. 4.4 – Analyse univariée des covariables

Variable	Valeur p du test de Wald
AGE	0.0005
SEXN	0.9913
SMOKEBLN	0.2599
BMIB	0.0496
FEV1B	<0.0001
TRT1PN	0.8691

la moins significative obtenue lors de l'analyse univariée, dans notre cas c'est **SEXN**. On regarde le pourcentage de changement provoqué par le retrait de cette variable par rapport au traitement. Ce pourcentage est de 19%, ce qui porte à croire que le sexe est confondant avec le traitement alors on gardera cette variable dans le modèle. Ensuite, on effectue la même démarche avec **SMOKEBLN** car c'est la variable suivante avec la moins forte significativité. On obtient qu'elle n'est ni confondante avec le traitement ni significative, donc on l'élimine du modèle. On fait de même avec toutes les autres covariables disponibles pour en arriver à la conclusion que l'âge et le volume expiré en une seconde sont significatives. Le modèle obtenu est donc le suivant :

$$\lambda(x) = \lambda_0(x)e^{0.32\beta_{SEXN}+0.03\beta_{AGE}-1.49\beta_{FEV1B}+0.06\beta_{TRT1PN}}$$

4.2.3 Forme fonctionnelle

À cette étape, on doit vérifier la forme fonctionnelle des deux variables continues du modèle : AGE et FEV1B. Par la méthode manuelle de lissage de courbe, on obtient les graphiques de la figure 4.1. Pour les deux variables, les courbes semblent plutôt droites.

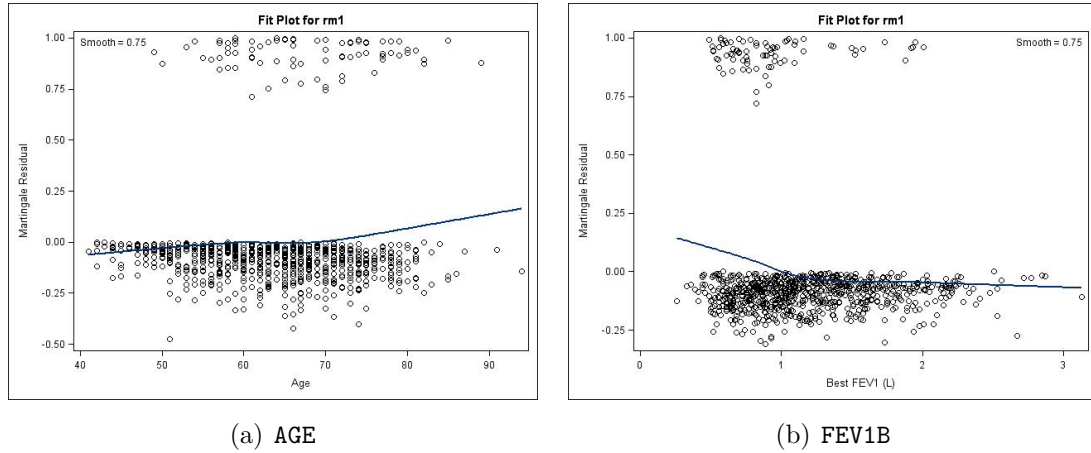


FIG. 4.1 – Forme fonctionnelle : Procédure LOESS

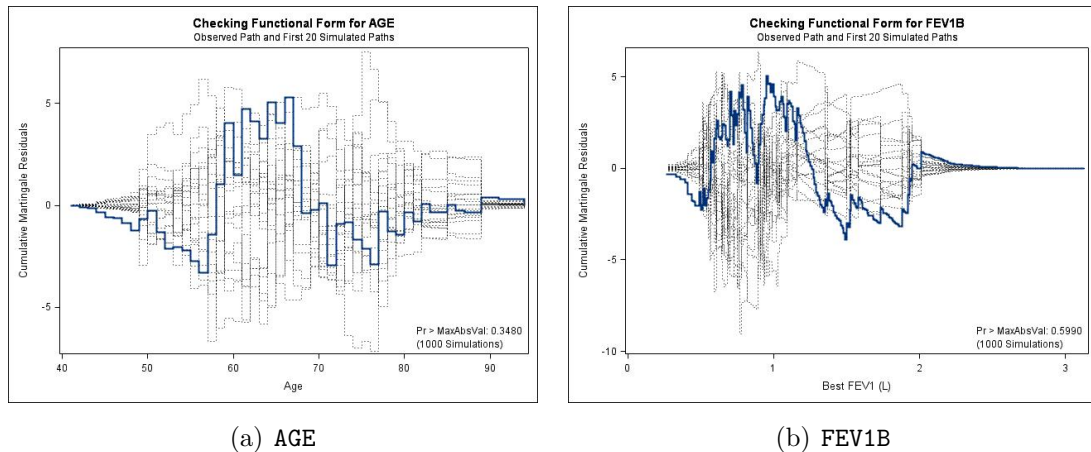


FIG. 4.2 – Forme fonctionnelle : Énoncé ASSESS dans la procédure PHREG. Valeurs p : AGE=0.3480 et FEV1B=0.5990

On peut donc penser que la forme continue est acceptable. Par contre, on observe une légère croissance après 70 ans et une légère décroissance avant un pour le volume expiré en une seconde. Est-ce que ces croissances et décroissances sont suffisantes pour exiger une modification des formes fonctionnelles ? Ce n'est pas évident à dire. Regardons les

résultats obtenus avec l'autre méthode.

En utilisant l'énoncé `ASSESS`, on obtient les graphiques et les valeurs p de la figure 4.2. Dans les deux cas, la valeur p est supérieure à 0.05 donc il n'y a pas d'évidence que la forme fonctionnelle ait mal été spécifiée. Contrairement à la méthode précédente, le test d'hypothèse nous permet de dire facilement si la forme fonctionnelle est bien spécifiée ou non. On peut donc conserver les variables sous leur forme continue.

4.2.4 Hypothèse de proportionnalité

Pour la méthode qui concerne les variables continues, on ajoute les variables $\text{AGEt}=\text{AGE}*\log(\text{TIM2PRD})$ et $\text{FEV1Bt}=\text{FEV1B}*\log(\text{TIM2PRD})$ au modèle. On obtient qu'aucune de ces nouvelles variables n'est significative puisque leurs valeurs p sont respectivement 0.8608 et 0.8273. Donc, on ne peut pas conclure que l'hypothèse de proportionnalité soit violée. Pour être plus rigoureux, il aurait fallu ajouter une à une ces variables au modèle. Mais comme les valeurs p obtenues sont loin d'être significatives, on s'est permis de le faire en une seule étape.

Si on regarde les graphiques du $\log(-\log(\hat{S}))$ en fonction du temps pour les deux variables discrètes dans la figure 4.3, on voit que pour `SEXN` les droites semblent parallèles. Pour ce qui est de `TRT1PN`, c'est moins évident de conclure que les droites sont parallèles. Ce qui nous intéresse, c'est de s'assurer qu'en un ou plusieurs points, les droites ne se croisent pas. Ici, on remarque qu'elles se croisent à plusieurs endroits. En fait, elles se croisent à tellement d'endroits qu'on arrive à la conclusion qu'elles ne sont tout simplement pas significativement différentes. À partir de cette observation, il est difficile de dire si elles se croisent. La prochaine méthode sera particulièrement utile pour déterminer si l'hypothèse de proportionnalité est respectée pour le traitement.

Pour chacune des covariables, l'option `PH` nous permet de vérifier facilement si l'hypothèse de proportionnalité est respectée. Les graphiques de la figure 4.4 nous montrent les résultats obtenus. On observe que toutes les valeurs p sont supérieures à 0.05, ce qui signifie que pour aucune covariable il n'y a d'évidence que l'hypothèse de proportionnalité soit violée. Pour la variable de traitement, il n'y a aucun problème de proportionnalité selon cette méthode alors qu'avec la deuxième méthode c'était difficile d'arriver à cette conclusion. Finalement, l'hypothèse de proportionnalité n'est pas violée pour aucune covariable, donc on peut utiliser le modèle de Cox pour ajuster nos données.

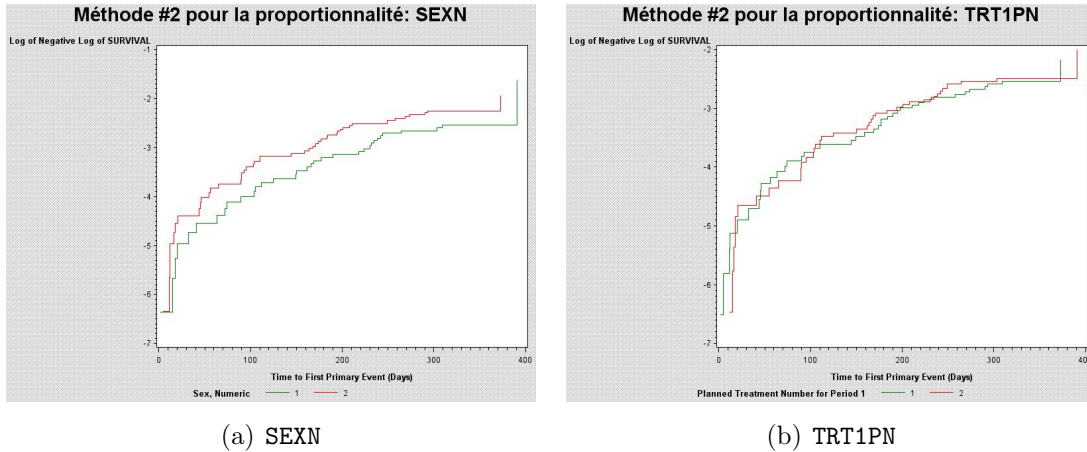


FIG. 4.3 – Hypothèse de proportionnalité : $\log(-\log(\hat{S}))$ en fonction du temps

4.2.5 Diagnostics d'influence (Valeurs aberrantes et données influentes)

Ici, on regarde ce qu'on obtient pour chacune des cinq statistiques d'influence. Pour les résidus de score standardisés, $DFBETAS$, on a fait sortir les 10 patients ayant les plus grands résidus pour chaque covariable continue. Ils sont présentés dans les tables 4.5 et 4.6. Selon notre taille d'échantillon, $n \approx 800$, seulement les patients avec $DFBETAS > 2\sqrt{\frac{1}{800}} \approx 0.07$ sont considérés potentiellement aberrants. Ce n'est le cas d'aucun patient.

La figure 4.5 montre les graphiques de LD et LMAX en fonction de la survie estimée. On remarque que pour les deux statistiques, un patient semble un peu à l'écart des autres patients avec les valeurs de $LD=0.1777$ et $LMAX=0.3520$. En regardant les données, on s'aperçoit que c'est le même patient (S029). Ce patient pourrait avoir une plus grande influence sur l'ajustement du modèle que les autres ; il faudra donc lui porter une attention particulière.

Pour les résidus de déviance, on a également fait sortir les 10 patients avec les plus grands résidus. Ils sont présentés dans la table 4.7. On remarque que 3 patients ont des résidus supérieurs à 3 ; il faudra également leur porter une attention particulière.

La figure 4.6 montre les résidus du score en fonction de chaque covariable continue. Aucun patient ne semble être extrême par rapport aux autres.

Au total, nous avons 4 différents patients potentiellement influents ou aberrants. Ils

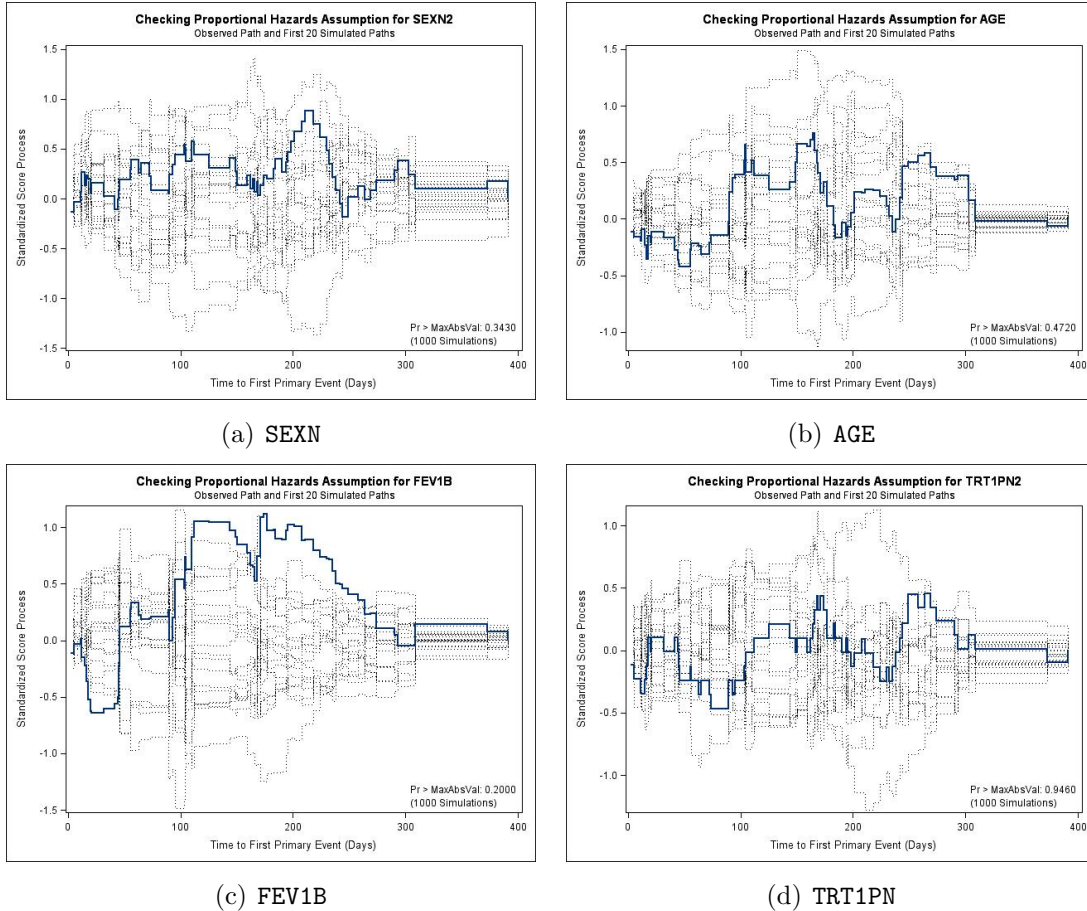


FIG. 4.4 – Hypothèse de proportionnalité : option PH dans ASSESS. Valeurs p : SEYN=0.3430, AGE=0.4720, FEV1B=0.2000 et TRT1PN=0.9460

représentent environ 0.5% de l'échantillon, ce qui n'est pas énorme. On a donc de bonnes raisons de croire que le modèle utilisé s'ajuste bien aux données.

4.2.6 Adéquation du modèle

Avant de regarder l'adéquation du modèle, commençons par vérifier que le modèle obtenu n'est pas biaisé. En ajoutant l'option `selection=score`, on obtient les C_p de Mallows de la table 4.8. Au total, on a six covariables donc 64 modèles possibles. On a fait imprimer les 3 meilleurs sous-ensembles de covariables pouvant former le modèle.

On se souvient que si $C_p \leq p$, le modèle n'est pas biaisé et toutes les variables importantes ont été incluses dans le modèle, aucune n'a été omise. Le modèle choisi, celui en gras dans la table 4.8, compte quatre variables et $C_p = 3.2715$ donc le modèle semble bon. Il est intéressant de mentionner que si on avait utilisé la méthode de sélection

TAB. 4.5 – Résidus de score standardisés, DFBETAS, pour AGE

USUBJID	TIM2PRMD	CENSPRM	SEXN	AGE	FEV1B	TRT1PN	dfAGE
S013	16	1	2	58	0.49	2	0.031867
S036	12	1	2	69	0.57	1	0.030101
S020	184	1	2	59	0.68	1	0.027216
S014	159	1	2	66	0.69	1	0.026143
S003	90	1	2	79	0.52	1	0.026056
S007	177	1	2	62	0.75	1	0.025704
S016	110	1	2	58	0.82	2	0.025028
S024	18	1	2	73	0.73	2	0.024972
S027	90	1	2	77	0.52	2	0.024961
S008	65	1	2	62	0.82	2	0.024683

TAB. 4.6 – Résidus de score standardisés, DFBETAS, pour FEV1B

USUBJID	TIM2PRMD	CENSPRM	SEXN	AGE	FEV1B	TRT1PN	dfFEV1B
S029	110	1	2	85	1.88	1	.003474296
S021	244	1	1	89	0.82	2	.002737692
S001	149	1	1	81	0.76	1	.002081858
S005	55	1	2	79	1.37	2	.002069980
S022	242	1	1	82	0.91	2	.002066105
S018	93	1	2	77	1.53	1	.001917224
S020	104	1	1	80	0.65	1	.001880579
S003	201	1	2	82	0.89	2	.001839344
S006	150	1	1	80	0.73	2	.001825308
S038	190	1	1	78	0.91	1	.001670765

des meilleurs sous-ensembles au lieu de la méthode manuelle, on aurait obtenu le même modèle. Selon les recommandations de Mallows, il faut prendre le premier modèle où C_p est plus petit ou égal au nombre de variables dans le modèle et pour lequel la variable d'intérêt est incluse. Dans notre cas, c'est la variable TRT1PN qui est très importante et qui doit faire partie du modèle. La figure 4.7 montre le C_p de Mallows en fonction du nombre de variables pour les meilleurs modèles.

Pour vérifier l'adéquation du modèle, on ajuste les deux modèles. Pour le modèle complet, on obtient les résultats de la table 4.9 et pour le modèle réduit, ceux de la table 4.10. À partir de ces résultats, on peut calculer la statistique du test du rapport des vraisemblances ($LRT=956.849-955.949=0.90071$), le nombre de degrés de liberté ($6-4=2$) et la valeur p de 0.6374. On peut donc conclure que le modèle réduit est adéquat.

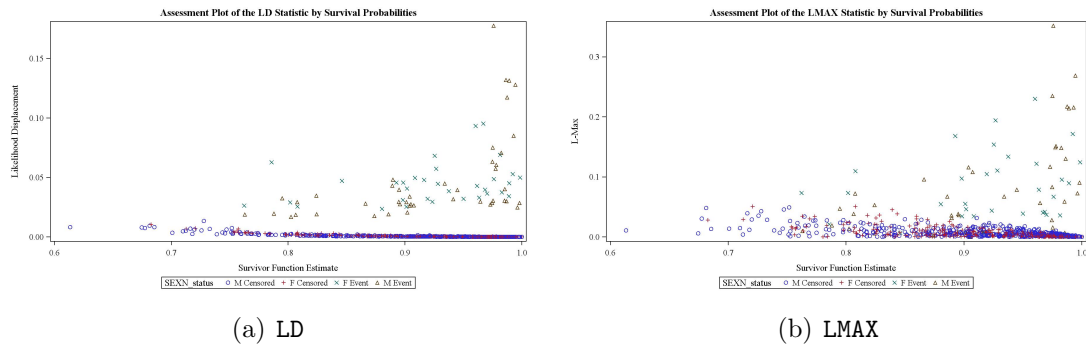


FIG. 4.5 – LD et LMAX

TAB. 4.7 – Résidus de déviance

USUBJID	TIM2PRMD	CENSPRM	SEXX	AGE	FEV1B	TRT1PN	deviance
S006	2	1	1	59	0.57	1	3.30332
S024	5	1	2	64	1.16	1	3.21813
S007	11	1	2	65	1.05	1	3.02627
S039	46	1	2	66	1.95	1	2.89648
S010	45	1	2	65	1.73	1	2.81566
S022	15	1	1	57	0.70	2	2.77748
S002	104	1	2	66	2.01	2	2.66074
S011	41	1	1	59	1.04	2	2.65267
S020	32	1	1	64	1.03	1	2.64645
S015	12	1	2	74	0.78	2	2.62427

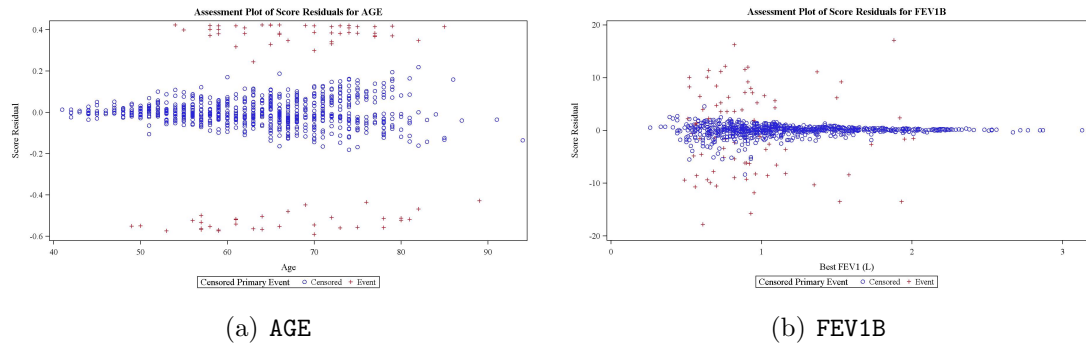


FIG. 4.6 – Résidus du score

TAB. 4.8 – Méthode de sélection des meilleurs sous-ensembles (C_p de Mallows)

Number InModel	mallows_ C_p	ScoreChiSq	scoreq	VariablesInModel
1	4.6253	23.4561	8.6253	FEV1B
1	16.0078	12.0735	20.0078	AGE
1	24.2226	3.8588	28.2226	BMIB
2	1.1836	28.8978	3.1836	AGE FEV1B
2	3.8862	26.1952	5.8862	SEXN2 FEV1B
2	4.3813	25.7001	6.3813	FEV1B BMIB
3	1.3529	30.7285	1.3529	SEXN2 AGE FEV1B
3	1.7298	30.3515	1.7298	AGE FEV1B BMIB
3	3.0680	29.0133	3.0680	SMOKEBLN1 AGE FEV1B
4	2.1195	31.9618	0.1195	SEXN2 AGE FEV1B BMIB
4	3.1519	30.9295	1.1519	SMOKEBLN1 SEXN2 AGE FEV1B
4	3.2715	30.8099	1.2715	TRT1PN2 SEXN2 AGE FEV1B
5	4.0528	32.0286	0.0528	TRT1PN2 SEXN2 AGE FEV1B BMIB
5	4.0657	32.0157	0.0657	SMOKEBLN1 SEXN2 AGE FEV1B BMIB
5	5.0740	31.0073	1.0740	TRT1PN2 SMOKEBLN1 SEXN2 AGE FEV1B
6	6.0000	32.0814	0.0000	TRT1PN2 SMOKEBLN1 SEXN2 AGE FEV1B BMIB

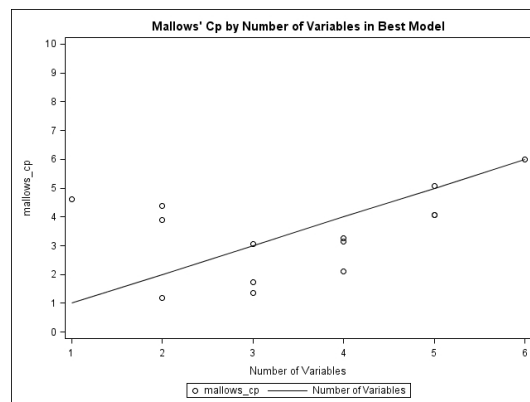


FIG. 4.7 – C_p de Mallows en fonction du nombre de variables

TAB. 4.9 – Ajustement du modèle complet
Model Fit Statistics

Criterion	Without Covariables	With Covariables
-2 LOG L	992.758	955.949
AIC	992.758	967.949
SBC	992.758	982.089

TAB. 4.10 – Ajustement du modèle réduit
Model Fit Statistics

Criterion	Without Covariables	With Covariables
-2 LOG L	992.758	956.849
AIC	992.758	964.849
SBC	992.758	974.276

4.3 Analyse principale

Maintenant que nous avons obtenu le modèle qui explique le mieux l'évènement principal et que toutes les hypothèses ont été vérifiées, nous pouvons procéder à l'analyse principale du projet. On rappelle que le modèle obtenu est le suivant :

$$\lambda(x) = \lambda_0(x)e^{0.32\beta_{SEXN}+0.03\beta_{AGE}-1.49\beta_{FEV1B}+0.06\beta_{TRT1PN}}$$

et les tests d'hypothèses sont les suivants :

$$\begin{array}{ll} 1. \text{ Test de non-infériorité} & 2. \text{ Test de supériorité} \\ H_0 : HR\left(\frac{T}{C}\right) \geq 1.4 & H_0 : HR\left(\frac{T}{C}\right) \geq 1 \\ H_1 : HR\left(\frac{T}{C}\right) < 1.4 & H_1 : HR\left(\frac{T}{C}\right) < 1 \end{array}$$

TAB. 4.11 – Résultat de l'analyse principale

The PHREG Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio	90% Hazard Ratio Confidence Limits
SEXN 2	1	0.31876	0.23596	1.8249	0.1767	1.375	0.933 2.028
AGE	1	0.02989	0.01301	5.2794	0.0216	1.030	1.009 1.053
FEV1B	1	-1.49808	0.33668	19.7981	<.0001	0.224	0.128 0.389
TRT1PN 2	1	0.06376	0.22684	0.0790	0.7786	1.066	0.734 1.548

Lorsqu'on a ajusté le modèle de Cox aux données obtenues lors de cet étude clinique, on a obtenu les résultats de la table 4.11. Le risque relatif pour la variable de traitement est de 1.066 avec l'intervalle de confiance [0.734,1.548]. Comme la borne supérieure est supérieure à 1.4, on ne peut pas démontrer la non-infériorité. On peut interpréter ce résultat comme une augmentation significative du nombre d'évènements principaux associé au traitement par rapport au contrôle, et donc qu'il y a possiblement un risque à utiliser le traitement. Ceci est le résultat obtenu avec les données bidons donc il est également bidon.

4.4 Analyses secondaires

4.4.1 Interaction entre les groupes de traitement et chaque covariable

Comme première analyse secondaire, on s'intéresse à l'interaction potentielle entre les groupes de traitement et chaque covariable prise une à une. Plus tôt, on a testé si on devait ajouter des termes d'interaction dans le modèle et on a conclu que non, car aucune n'était significative. Alors pourquoi faire cette analyse rendu à ce point ? La réponse est que, même si les groupes de **SEXN** n'ont pas d'effet significativement différent sur la survie, le traitement pourrait avoir un effet différent sur l'un d'eux.

Regardons justement la variable **SEXN**. L'ajustement du modèle de Cox auquel on a

TAB. 4.12 – Interaction entre TRT1PN et SEXN

Hazard Ratios for TRT1PN			
Description	Point	90% Wald Confidence	
	Estimate	Limits	
TRT1PN 1 vs 2 At SEXN=1	0.713	0.401	1.269
TRT1PN 1 vs 2 At SEXN=2	1.146	0.701	1.875

TAB. 4.13 – Interaction entre TRT1PN et AGE

Hazard Ratios for TRT1PN			
Description	Point	90% Wald Confidence	
	Estimate	Limits	
TRT1PN 1 vs 2 At AGE=50	0.719	0.312	1.657
TRT1PN 1 vs 2 At AGE=60	0.836	0.511	1.370
TRT1PN 1 vs 2 At AGE=70	0.974	0.661	1.434

ajouté **TRT1PN*SEXN** nous donne le risque relatif de **TRT1PN** pour chacun des groupes de **SEXN**. D'après la table 4.12, on peut voir que pour les femmes (**SEXN=1**), la borne supérieure de l'IC à 90% est 1.269. Contrairement au résultat obtenu à l'analyse principale, on peut conclure que pour les femmes, le traitement est non-inférieur au contrôle, donc aucun risque excessif à utiliser le traitement. Comme mentionné au tout début, le but des analyses secondaires n'est pas d'apporter des conclusions éprouvées, mais plutôt des pistes à de futures études cliniques. Par exemple, l'objectif principal d'une

prochaine étude clinique pourrait être de vérifier si le traitement est non-inférieur au contrôle pour certains groupes comme les femmes. De façon similaire, on obtient que pour les personnes de 60 ans, il n'y aurait aucun risque excessif à utiliser le traitement.

4.4.2 Fonction de survie

Grâce à l'estimateur de Kaplan-Meier, les fonctions de survie sont estimées pour les strates TRT1PN et SEXN. On peut voir l'allure générale des fonctions obtenues dans les graphiques 4.8. Selon le test du log-rang, on obtient une valeur p de 0.8694 pour

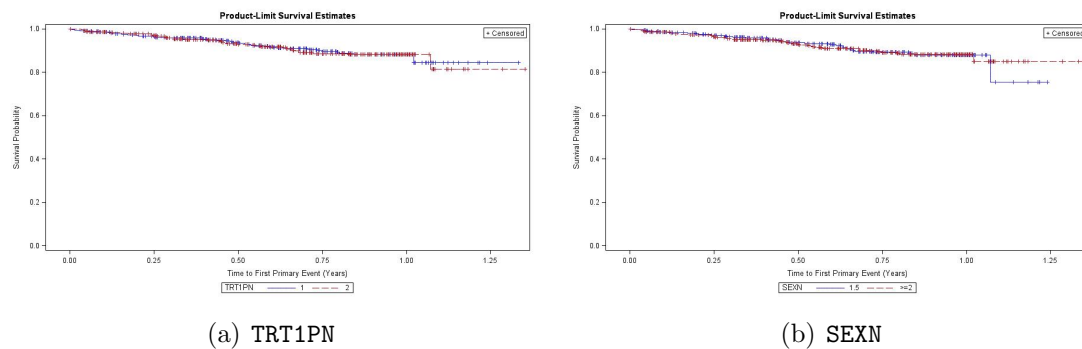


FIG. 4.8 – Fonction de survie

TRT1PN et 0.9917 pour SEXN. Dans les deux cas, il n'y a pas de différences significatives entre les strates étudiées, donc les strates semblent homogènes.

Chapitre 5

Discussion et conclusion

Le but de cette étude était d'évaluer la efficacité et l'innocuité de l'utilisation à long terme d'un médicament dans le traitement de la MPOC. Plus particulièrement, de déterminer si l'utilisation de ce médicament était non-inférieure à celle d'un médicament déjà reconnu comme traitement de la MPOC. Pour ce faire on a ajusté un modèle de Cox aux données recueillies lors d'une étude clinique réalisée chez Quintiles Canada Inc.. Parmi les variables disponibles au départ, seules l'âge, le sexe et le volume expiré en une seconde se sont avérées significatives ou confondantes avec le traitement. Une fois le modèle choisi, on a regardé que les variables continues ne devaient pas être modifiées et que leur forme continue était appropriée. Ensuite, on a vérifié que l'hypothèse de proportionnalité était respectée pour toutes les covariables, ce qui était le cas. On a également vérifié que le modèle choisi s'ajustait bien aux données à l'aide des diagnostics d'influence, d'un indicateur de biais et de l'adéquation du modèle. Finalement, on a répondu à l'objectif principal en regardant la borne supérieure de l'intervalle de confiance obtenu pour le risque relatif du traitement. Avec les données disponibles, on en est arrivé à la conclusion qu'il y a une augmentation significative du nombre d'événements principaux pour les gens qui ont pris le médicament à l'étude par rapport à ceux qui ont pris le contrôle. Donc, si on se réfère à l'hypothèse initiale, l'étude ne serait pas concluante.

Il est à noter que l'étude n'était pas encore terminée au moment de la rédaction de cet essai, car le nombre d'événements principaux de 86 n'avait pas encore été observé. Ceci pourrait avoir certains effets sur l'analyse finale car les résultats pourraient s'avérer quelque peu différents à la toute fin de l'étude par rapport à ceux qu'on a utilisés. Parmi les résultats obtenus au moment de cette analyse, on a remarqué un nombre élevé de censures et un faible nombre d'événements observés. On rappelle que si le nombre d'événements observé est plus faible que ce qui était planifié au départ, il pourrait en

résulter une diminution de la puissance de l'étude. Pour s'assurer que ce ne sera pas le cas à la fin de l'étude, il faudra attendre les 86 événements avant de terminer l'étude. Un autre point important à mentionner à ce moment concerne l'aveuglement. Les données avec lesquelles nous avons travaillé étaient incomplètes dans le sens où la variable de traitement n'était pas réelle. Elle a été aléatoirement modifiée afin de préserver l'aveuglement. On peut donc supposer que les résultats obtenus lors des analyses effectuées sur la variable de traitement sont bidons autant que la variable elle-même. Par contre, une fois que l'étude sera terminée, toute la démarche effectuée dans cet essai pourra être reproduite et, à ce moment, on obtiendra les vrais résultats.

Regardons ce qui a été fait afin de s'assurer d'avoir de résultats robustes et fiables. Outre la puissance, il y a d'autres facteurs que je n'ai pas mentionnés qui ont été pris en considération lors de la détermination de la taille d'échantillon tels la proportion annuelle des événements et des patients perdus de vue. Le fait de prendre ces facteurs en considération permet de prendre un échantillon suffisamment grand et d'assurer des résultats fiables. Ensuite, toutes les vérifications qui ont été faites concernant le modèle de Cox (forme fonctionnelle, hypothèse de proportionnalité, diagnostic d'influence et adéquation du modèle) nous permettent encore une fois de croire que nos résultats sont robustes et fiables.

Finalement, il est un peu dommage d'avoir réalisé toutes ces analyses sur des données bidons car, dépendamment des vrais résultats que l'on obtiendra avec les données finales, il sera plus facile d'apporter des pistes de futures études. Pour l'instant, nos résultats sur l'ensemble des données ne sont pas concluants. Par contre, il y aurait peut-être du potentiel à étudier ce médicament seulement sur les femmes ou sur le groupe d'âge de 60 ans.

Chapitre 6

Annexes

6.1 Programmes SAS

Cette annexe contient tous les code SAS utilisés lors des analyses présentées dans cet essai-stage.

Section 3.2.1 Estimation de la fonction de survie grâce à l'estimateur de Kaplan-Meier

```
proc lifetest data=adae2 plots=(lls) ;  
strata TRT1PN ;  
time TIME2PRM*CENSPRM(0) ;  
title "Analyse 3a : Fonction de survie et incidence cumulative par strate  
TRT1PN" ;  
run ;
```

Section 3.3 Ajustement du modèle semi-paramétrique de Cox

```
proc phreg data=adae2 alpha=0.01 ;  
class TRT1PN(ref='1') SMOKEBLN SEXN(ref='1') ;  
model TIM2PRMD*CENSPRM(0) = TRT1PN SMOKEBLN SEXN AGE BMIB FEV1B / ties=efron ;  
run ;
```

Section 3.3.4 Vérification de la forme fonctionnelle selon la première méthode : LOESS

```
proc phreg data=adae2 noprint ;  
model TIM2PRMD*CENSPRM(0) = FEV1B SEXN TRT1PN/ties=efron ;
```

```

output out=resm resmart=rm1 ;
title "Méthode #2 pour vérifier la forme fonctionnelle des variables continues :
Courbe de lissage - AGE" ;
run ;

data residu ;
merge adae2 resm ;
by USUBJID ;
run ;

proc loess data=residu plots(unpack) ;
model rm1=AGE/smooth=0.75 ;
title "Méthode #2 pour vérifier la forme fonctionnelle des variables continues :
Courbe de lissage - AGE" ;
run ;

```

Vérification de la forme fonctionnelle selon la seconde méthode : ASSESS

```

proc phreg data=adae2 alpha=0.01 ;
class TRT1PN(ref='1') SEXN(ref='1') ;
model TIM2PRMD*CENSPRM(0) = TRT1PN SEXN AGE FEV1B / ties=efron ;
assess var=(AGE FEV1B) / resample seed=54321 ;
run ;

```

Section 3.3.5 Vérification des hypothèses de proportionnalité selon la première méthode : Interaction des variables continues avec le temps.

```

proc phreg data=adae2 ;
model TIM2PRMD*CENSPRM(0) = AGE FEV1B AGEt FEV1Bt ;
AGEt=AGE*log(TIM2PRMD) ;
FEV1Bt=FEV1B*log(TIM2PRMD) ;
proportionality_test : test AGEt, FEV1Bt ;
title "Méthode #2 pour vérifier la proportionnalité" ;
run ;

```

Vérification des hypothèses de proportionnalité selon la seconde méthode : Graphique du $\log(-\log(\hat{S}))$ en fonction du temps.

```

proc phreg data=adae2 ;
class TRT1PN(ref='1') ;

```

```

model TIM2PRMD*CENSPRM(0) = TRT1PN AGE FEV1B / ties=efron;
strata SEXN;
baseline out=a loglogs=lls survival=s;
title "Méthode #2 pour la proportionnalité : SEXN";
run;

```

```

proc gplot data=a;
plot lls*TIM2PRMD=SEXN;
symbol1 v=none i=stepJ c=green line=1;
symbol2 v=none i=stepJ c=red line=1;
run;

```

Vérification des hypothèses de proportionnalité selon la troisième méthode : PH.

```

proc phreg data=adae2 alpha=0.01;
class TRT1PN(ref='1') SEXN(ref='1');
model TIM2PRMD*CENSPRM(0) = TRT1PN SEXN AGE FEV1B / ties=efron;
assess var=(AGE FEV1B) PH / resample seed=54321;
run;

```

Section 3.3.6 Vérification des valeurs aberrantes et données influentes

```

proc phreg data=adae2 noprint;
class TRT1PN(ref='1') SEXN(ref='1');
model TIM2PRMD*CENSPRM(0) = SEXN AGE FEV1B TRT1PN /ties=efron;
output out=output.resid resdev=deviance ressko=scoreAGE scoreFEV1B
ld=likedisp lmax=maxdisp survival=surv dfbeta=dfAGE dfFEV1B;
run;

```

```

proc format;
value censor 1='Event'
0='Censored';
run;

```

```

ods graphics on / width = 9in height = 5in noborder;

```

```

proc sgplot data=output.resid;
scatter y=scoreAGE x=AGE / group=CENSPRM;
format CENSPRM censor.;
yaxis label='Score Residual';

```

```
title 'Assessment Plot of Score Residuals for AGE';
run;
quit;

data output.resid;
set output.resid;
SEXN_status=compress(SEXN||CENSPRM);
run;

proc format;
value $censor 10='F Censored'
11='F Event'
20='M Censored'
21='M Event';
run;

proc sgplot data=output.resid;
scatter y=maxdisp x=surv / group=SEXN_status;
yaxis label='L-Max';
format SEXN_status $censor.;
title 'Assessment Plot of the LMAX Statistic by Survival Probabilities';
run;
quit;

proc sgplot data=output.resid;
scatter y=likedisp x=surv / group=SEXN_status;
yaxis label='Likelihood Displacement';
format SEXN_status $censor.;
title 'Assessment Plot of the LD Statistic by Survival Probabilities';
run;
quit;

proc rank data=output.resid groups=100 out=ranks;
var likedisp maxdisp;
ranks likegroup maxgroup;
run;

proc sort data=ranks;
by descending maxdisp;
run;
```

```

proc print data=ranks ;
where likegroup = 99 or maxgroup=99 ;
var TIM2PRMD CENSPRM AGE FEV1B scoreAGE scoreFEV1B deviance
likedisp maxdisp surv dfAGE dfFEV1B likegroup maxgroup ;
id maxdisp ;
title 'Observations with Possible Outlying Covariate Patterns' ;
run ;

```

Section 3.3.7 Vérification de l'adéquation du modèle selon la première méthode : C_p de Mallows

```

proc phreg data=adae2 alpha=0.01 ;
class TRT1PN(ref='1') SEXN(ref='1') SMOKEBLN ;
model TIM2PRMD*CENSPRM(0) = TRT1PN SMOKEBLN SEXN AGE FEV1B BMIB/ selection=score
best=3 ties=efron rl ;
title "méthode de sélection des meilleurs sous-ensembles" ;
ods output bestsubsets=score ;
run ;

```

```

proc print data=score(drop=control_var obs=12) ;
title "Line Listing of Score Data Set" ;
run ;

```

```

data _null_ ;
set score ;
call symput ('ChisqFullModel',scorechisq) ;
call symput ('ParmsFullModel',numberinmodel) ;
run ;

```

```

data score ;
set score ;
scoreq=&ChisqFullModel-scorechisq ;
q=&ParmsFullModel-numberinmodel ;
mallows_cp=scoreq+(&ParmsFullModel-(2*q)) ;
run ;

```

```

proc print data=score noobs ;
var numberinmodel mallows_cp scorechisq scoreq variablesinmodel ;
title "Mallows' Cp by Number of Variables in Model" ;
run ;

```

```

proc sgplot data=score ;
scatter y=mallows_cp x=numberinmodel ;
series y=numberinmodel x=numberinmodel ;
yaxis values=(0 to 10 by 1) ;
xaxis values=(1 to 6 by 1) ;
title "Mallows' Cp by Number of Variables in Best Model" ;
run ;
quit ;

```

Vérification de l'adéquation du modèle selon la seconde méthode : Test du rapport des vraisemblances

```

ods output globaltests(match_all persist=proc)=test ;
proc phreg data=adae2 ;
class TRT1PN(ref='1') SEXN(ref='1') SMOKEBLN ;
model TIM2PRMD*CENSPRM(0) = TRT1PN SMOKEBLN SEXN AGE FEV1B BMIB / ties=efron ;
run ;
proc phreg data=adae2 ;
class TRT1PN(ref='1') SEXN(ref='1') ;
model TIM2PRMD*CENSPRM(0) = TRT1PN SEXN AGE FEV1B / ties=efron ;
run ;

data report ;
merge test(rename=(chisq=chisq1 df=df1)) test1(rename=(chisq=chisq2 df=df2)) ;
if test = 'Likelihood Ratio' ;
df=abs(df2-df1) ;
LRT=abs(chisq2-chisq1) ;
pvalue=1-probchi(LRT,df) ;
run ;

proc print data=report split='*' noobs ;
var LRT df pvalue ;
label LRT='likelihood ratio*test statistic*comparing*model2 and model1'
df='degrees of*freedom'
pvalue='p-value' ;
title 'Test d'adéquation globale' ;
run ;

```

Bibliographie

- [1] Agence de la santé publique du Canada. *Faits et chiffres sur les maladies respiratoires chroniques*.
Site internet : http://www.phac-aspc.gc.ca/cd-mc/crd-mrc/mrc_figures-crd_figures-fra.php#1-8
- [2] Association pulmonaire du Canada. *MPOC*.
Site internet : http://www.poumon.ca/diseases-maladies/copd-mpoc/what-quoi/index_f.php
- [4] Belsey, D. A., Kuh, E. et Welsch, R. E. 1980. *Regression Diagnostics : Identifying Influential Data and Sources of Collinearity*. New York : John Wiley & Sons.
- [4] Klein, John P., Moeschberger, Melvin L. 2003. *Survival Analysis Techniques for Censored and Truncated Data*. Springer-Verlag New York, Inc. 2nd edition.
- [5] Lin, D. Y., Wei, L. J. and Ying, Z. *Checking the Cox Model with Cumulative Sums of Martingale-Based Residuals*. *Biometrika*, 1993, 80, pp.557-572.
- [6] Mallows, C. L. *Some Comments on C_p* . *Technometrics*, 1973, Vol. 15, pp. 661-675.
- [7] Samson, Élodie. 2008 *Identification des marqueurs pronostiques chez les patients atteints d'un cancer de la tête et du cou*.
- [8] Santé Canada. *Innocuité et efficacité des médicaments génériques*.
Site internet : <http://www.hc-sc.gc.ca/hl-vs/iyh-vsv/med/med-gen-fra.php>
- [9] Santé Canada. *Ligne directrice - E10*.
Site internet : http://www.hc-sc.gc.ca/dhp-mps/prodpharma/applic-demande/guide-ld/ich/efficac/e10_step4-fra.php#a2.4
- [10] SAS Institute Inc. *Statistics 1 : Introduction to ANOVA, Regression, and Logistic Regression*. SAS Course Notes, 2011
- [11] SAS Institute Inc. *Survival Analysis Using the Proportional Hazards Model*. SAS Course Notes, 2009

- [12] Sherwood, Lauralee. *Physiologie humaine*. 2e édition, Édition de boeck Supérieur, 2006
- [13] U.S. Department of Health and Human Services, Food and Drug Administration (FDA). *Guidance for Industry Non-Inferiority Clinical Trials (March 2010)*.
Site internet : <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM202140.pdf>