

Note pédagogique

Variables dépendantes du temps dans le modèle de Cox Théorie et pratique

Time-dependent covariates in the Cox proportional hazards model. Theory and practice

L. DESQUILBET, L. MEYER

INSERM-INED U569, Service d'Épidémiologie, Hôpital de Bicêtre, 82, rue du Général-Leclerc, 94276 Le Kremlin-Bicêtre.
Email : desquilbet@vjf.inserm.fr (Tirés à part : L. Desquilbet)

In survival analysis, exposure that appears or changes during the follow-up of subjects must be taken into account as a time-dependent covariate in the Cox proportional hazards model. Two types of time-dependent covariates are defined: covariates with unique change, and covariates with multiple changes. The way of taking into account such changes in the exposure is presented in theory and illustrated from a small sample of 5 subjects enrolled in a French HIV cohort. The problems raised by missing data as well as alternative but more sophisticated solutions are also evocated. Annexes include programs for SAS software, and results from the "Log" and "Results" windows.

Survival analysis. Cox proportional hazards model. Time-dependent covariate. SAS software.

En analyse de survie, une exposition qui varie au cours du suivi de sujets doit être prise en compte en tant que variable dépendante du temps dans le modèle de Cox. Deux types de variables dépendantes du temps sont définis : les variables à changement unique et les variables à changements multiples. La façon dont ces deux types de variables sont pris en compte est présentée en théorie, puis illustrée à l'aide d'un exemple de 5 sujets issus d'une cohorte française de patients séropositifs. Les problèmes posés par les données manquantes ainsi que les solutions alternatives plus sophistiquées sont également évoquées. En annexes figurent les fichiers de données, la programmation commentée des variables dépendantes du temps avec le logiciel SAS, ainsi que les différents résultats qui apparaissent dans les fenêtres du logiciel.

Analyse de survie. Modèle de Cox. Variables dépendantes du temps. Logiciel SAS.

INTRODUCTION

Le modèle de Cox [1] est l'un des modèles les plus fréquemment utilisés pour les analyses de survie dans le domaine biomédical. Ce modèle semi-paramétrique permet d'étudier le risque de survenue d'un événement en tenant compte de facteurs de confusion potentiels et de quantifier, grâce au risque relatif (rapport des risques instantanés), l'association entre une exposition et le risque de survenue de l'événement étudié. Une des

hypothèses sur lesquelles ce modèle repose est celle de la proportionnalité des risques entre les groupes comparés. Des tests permettent de vérifier l'acceptabilité de cette hypothèse [2-4].

Dans les enquêtes de cohortes, les expositions dont on veut étudier l'effet sont classiquement des variables fixes, définies à la date d'origine du suivi (t_0). Elles concernent les caractéristiques intrinsèques des sujets (par exemple, le sexe ou le niveau d'études), les expositions passées (par

exemple, les antécédents de pathologies), ou les expositions présentes à t_0 (par exemple, la consommation de tabac). Il peut être intéressant, lorsque le protocole le permet, d'étudier l'effet d'une exposition susceptible de survenir ou de varier après t_0 . La prise en compte d'une telle exposition en tant que variable « fixe » à t_0 peut conduire à des erreurs d'interprétation, comme par exemple dans l'analyse des données du Stanford Heart Transplantation Program [5]. Les auteurs avaient étudié en analyse de survie l'effet de la survenue d'une transplantation cardiaque au cours du suivi sur le risque de décès en l'introduisant comme variable « fixe ». L'effet très protecteur de la transplantation cardiaque trouvé par les auteurs a été remis en cause par Gail qui a invoqué un biais de sélection majeur : les patients en « bonne » santé à l'inclusion dans l'étude pouvaient davantage bénéficier d'une transplantation cardiaque que les autres, ces derniers décédant plus rapidement [6]. Par la suite, Crowley et Hu ont repris les analyses avec un modèle de Cox en prenant en compte la transplantation cardiaque en tant que variable dépendante du temps, et ont confirmé les critiques énoncées par Gail [7].

Dans la littérature, on trouve de multiples exemples de traitement de variables dépendantes du temps : on peut citer l'étude du rôle de la transplantation du foie dans la progression d'une maladie inflammatoire de l'intestin [8], ou de celui de l'albumine sérique, du taux de CD4, et de la charge virale comme marqueur de progression de la maladie VIH [9, 10]. Bien que la théorie des variables dépendantes du temps ait déjà été décrite dans la littérature [11, 12], la manière dont elles sont prises en compte dans le modèle de Cox est rarement abordée.

Cet article a pour objectif d'explicitier de façon pédagogique le traitement des variables dépendantes du temps dans les logiciels statistiques courants. Après un bref rappel concernant le modèle de Cox et la définition des variables dépendantes du temps à changements unique et multiples, nous utiliserons un exemple issu d'une étude épidémiologique. Le problème des données manquantes sera ensuite abordé, pour terminer par un dernier chapitre concernant les limites de l'utilisation des variables dépendantes du temps et la présentation de solutions alternatives.

Les annexes comportent les fichiers de données de l'exemple, les programmes commentés et réalisés avec le logiciel SAS (version 8.2 ; SAS Institute, Inc., Cary, North Carolina), ainsi que les résultats qui apparaissent dans les fenêtres « Journal » et « Sortie » du logiciel SAS. D'autres logiciels tels que Stata ou S-Plus sont capables de traiter des expositions en tant que variables dépendantes du temps et donnent les mêmes résultats que SAS [11], mais semblent actuellement moins répandus parmi les épidémiologistes francophones. Notons également que le logiciel R (logiciel « open-source », dérivé du logiciel S-Plus, librement accessible à partir du site Internet <http://www.r-project.org/>) est aussi capable de traiter des expositions en tant que variables dépendantes du temps.

QUELQUES RAPPELS À PROPOS DU MODÈLE DE COX

Le modèle de Cox est un modèle semi-paramétrique qui exprime le risque instantané de survenue d'un événement en fonction du temps : $h(t) = \lambda_0(t) \exp(\beta X)$ où $\lambda_0(t)$ est une fonction de risque de base, X un vecteur colonne de p variables, et β le vecteur ligne associé de p paramètres inconnus. Le principe du calcul de l'estimation du vecteur β a été largement décrit dans la littérature [13]. Rappelons qu'il réside dans la maximisation de la vraisemblance partielle, produit des vraisemblances V_j calculées pour chaque événement j s'exprimant de la façon suivante :

$$V_j = \frac{h_i(t_j)}{h_i(t_j) + h_{i+1}(t_j) + \dots + h_n(t_j)}$$

où t_j est le temps de survenue du $j^{\text{ème}}$ événement, i l'indice du sujet qui présente l'événement au temps t_j , et $h_i(t_j)$ le risque instantané pour le sujet i au temps t_j . Les sujets d'indices i jusqu'à n sont les sujets considérés comme à risque de présenter l'événement à t_j . Il est important de remarquer dans l'expression de V_j que seuls les temps de survie des sujets ayant présenté l'événement sont pris en compte pour estimer le vecteur β .

Lorsque deux sujets présentent l'événement au même temps (ex-aequo), trois principales méthodes de maximisation de la vraisemblance partielle sont proposées par les logiciels : l'approximation de Breslow [14] (méthode utilisée par défaut pour

les logiciels SAS et Stata), l'approximation d'Efron [15] (méthode utilisée par défaut pour les logiciels R et S-Plus), ainsi qu'une méthode de calcul exact de la vraisemblance partielle [16]. La procédure SAS à utiliser pour analyser les données de survie avec le modèle de Cox est la procédure PROC PHREG, décrite en *Annexe 1*.

LES VARIABLES DÉPENDANTES DU TEMPS

DÉFINITION GÉNÉRALE

Une variable dépendante du temps dans le modèle de Cox est une variable qui prend en compte la modification de l'exposition d'un sujet au cours de son suivi dans une enquête de cohorte : si $X_i(t)$ est l'exposition du sujet i à un temps t , le risque instantané pour le sujet i au temps t s'écrit $h_i(t) = \lambda_0(t) \exp(\beta X_i(t))$. Le rapport des risques instantanés $h_i(t)$ et $h_j(t)$ de deux sujets i et j dont l'exposition respective au temps t vaut $X_i(t)$ et $X_j(t)$:

$$\frac{h_i(t)}{h_j(t)} = \exp(\beta(X_i(t) - X_j(t)))$$

ne peut plus être supposé proportionnel, dans la mesure où il dépend du temps. Cependant, l'interprétation du risque relatif reste valable [12, 17].

La variable dépendante du temps a un statut très particulier dans le modèle, dans la mesure où la valeur de l'exposition d'un même sujet qui sera prise en compte dans le calcul du risque relatif varie aux différents temps d'événements considérés. En effet, l'exposition d'un sujet au moment de la survenue du $k^{\text{ème}}$ événement pourra être différente au moment de la survenue du $(k+1)^{\text{ème}}$ événement. Une façon simple de considérer l'exposition est de prendre celle la plus proche précédant l'événement considéré. D'autres prises en compte de l'exposition telles que son cumul ou sa moyenne jusqu'à l'événement considéré peuvent également être utilisées. À partir de cette définition générale, deux types de variables dépendantes du temps peuvent être définis : les variables à changement unique et les variables à changements multiples.

LES VARIABLES DÉPENDANTES DU TEMPS À CHANGEMENT UNIQUE

Les variables dépendantes du temps à changement unique sont relatives à des expositions dont

le codage est binaire, et ne peuvent changer de statut (de « non exposé » vers « exposé », ou plus rarement d'« exposé » vers « non exposé ») qu'une seule fois au cours du suivi. Le codage en 0/1 permet une interprétation facile des risques relatifs, dans la mesure où l'exponentielle du paramètre β est une estimation du risque relatif des sujets exposés par rapport aux sujets non exposés.

Les expositions qui peuvent être traitées en tant que variables à changement unique sont des expositions irréversibles (par exemple, avoir bénéficié d'une transplantation cardiaque), ou des expositions réversibles dont on veut étudier l'effet de la première occurrence (par exemple, la survenue de troubles métaboliques). Trois cas de figure peuvent se présenter : a) un sujet est exposé à la date d'origine du suivi, il le reste par la suite ; b) un sujet n'est pas exposé à la date d'origine du suivi, devient exposé au cours de son suivi et le reste ; c) un sujet n'est jamais exposé au cours de son suivi. Si le statut vis-à-vis de l'exposition change plus d'une fois au cours du suivi et que l'on veut en tenir compte dans les analyses, l'exposition devra être prise en compte en tant que variable dépendante du temps à changements multiples.

La date à laquelle le sujet devient exposé doit être disponible, elle est connue précisément ou estimée. Dans les analyses, un sujet est considéré comme exposé au temps de l'événement considéré si son exposition survient avant cet événement. L'hypothèse nulle (H_0) que l'on teste est la suivante : le fait de devenir exposé au cours du suivi n'est pas associé au risque de survenue de l'événement par la suite.

VARIABLES DÉPENDANTES DU TEMPS À CHANGEMENTS MULTIPLES

Les variables dépendantes du temps à changements multiples sont relatives à des expositions qui sont recueillies de façon régulière dans l'étude (par exemple, la pression artérielle systolique, ou l'observance à un médicament qui peut être bonne, partielle ou nulle). Leur codage pourra être quantitatif, qualitatif à plusieurs classes, ou binaire (par exemple, la prise d'un traitement en 0, 1, 0, 1, etc. dans le cas d'arrêts et de reprises). En pratique, la valeur de l'exposition

prise en compte dans les comparaisons est souvent celle de la dernière mesure réalisée avant l'événement considéré. L'hypothèse nulle (H_0) que l'on teste est la suivante : la valeur de l'exposition recueillie au cours du suivi n'est pas associée au risque de survenue de l'événement par la suite.

ILLUSTRATION DES VARIABLES DÉPENDANTES DU TEMPS À L'AIDE D'UN EXEMPLE

PRÉSENTATION DE L'EXEMPLE

L'exemple choisi pour illustrer la prise en compte d'une exposition qui varie au cours du suivi est issu d'une étude portant sur des sujets récemment infectés par le virus de l'immunodéficience humaine (VIH) inclus dans la cohorte PRIMO [18]. L'un des objectifs de l'étude était de rechercher les facteurs associés à l'usage du préservatif lors de relations sexuelles rapportées par le sujet. Le comportement sexuel était recueilli à chaque visite, et concernait l'usage du préservatif entre la précédente consultation et la consultation courante. Les variables étudiées concernaient des expositions recueillies à l'inclusion, et pour certaines réactualisées au cours du suivi. Cette étude a montré que la survenue d'une lipodystrophie¹ au cours du suivi était liée à la survenue ultérieure d'une relation sexuelle sans préservatif. En revanche, elle n'a pas mis en évidence de lien entre le niveau d'immunodépression du sujet (représenté par le nombre de CD4 par mm^3 recueilli à chaque consultation) et l'usage du préservatif. Parmi les 223 sujets que comptait l'étude, nous en avons sélectionné 5. L'événement étudié est la première relation sexuelle sans préservatif (3 événements et 2 censures parmi les 5 sujets). Pour simplifier, nous prendrons ici comme date d'événement, la date de la visite à laquelle la relation sexuelle sans préservatif était déclarée. L'exposition traitée en tant que variable à changement unique est la lipodystrophie (paragraphe 4.2) ; celle traitée en tant que variable à changements multiples est le taux de CD4 recueilli à chaque visite (paragraphe 4.3), codée de manière quantitative en nombre de CD4 par

mm^3 . Les hypothèses testées sont les suivantes : a) la survenue d'une lipodystrophie au cours du suivi n'est pas associée à la survenue d'une relation sexuelle sans préservatif par la suite ; b) le taux de CD4 recueilli à la visite $n^{\circ}i$ n'est pas associé à la survenue d'une relation sexuelle sans préservatif déclarée à la visite $n^{\circ}i+1$.

Dans un but pédagogique, nous avons testé les hypothèses dans deux fichiers de données décrits en *Annexe 2* : un fichier de données initial (configuration 1) et un fichier modifié (configuration 2). La modification concerne seulement le sujet C. Les figures 1 et 2 illustrent ces deux fichiers de données. Selon le principe du calcul des vraisemblances V_j , seuls les temps de survie des sujets qui présentent l'événement (sujets A, C et D) sont considérés.

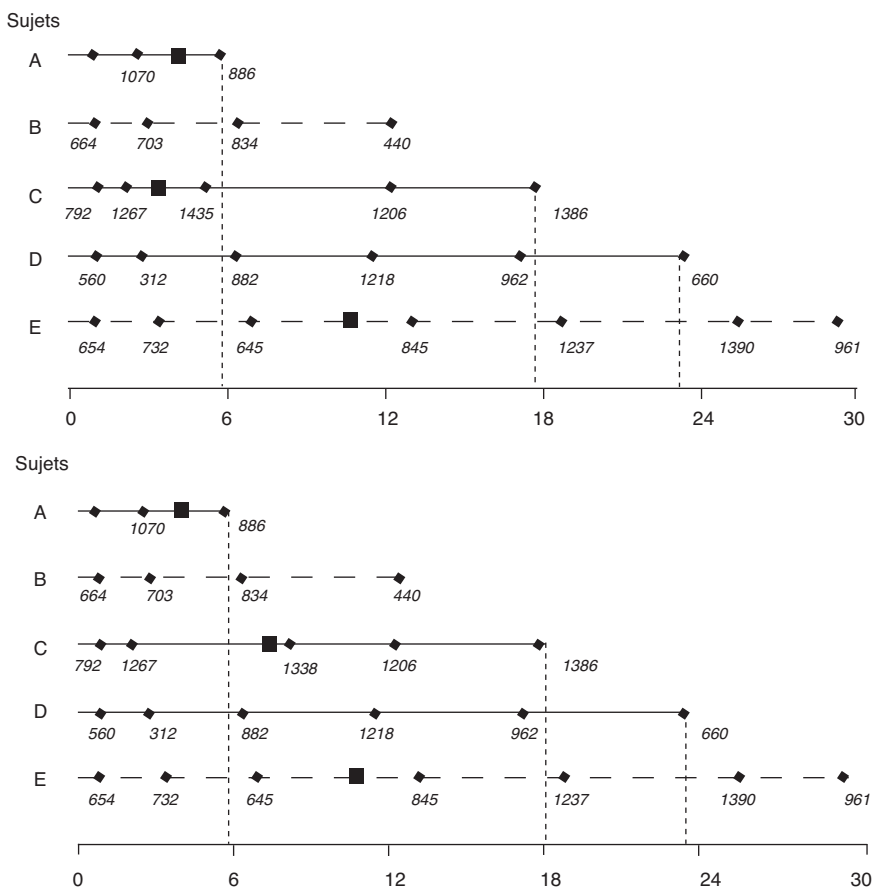
ILLUSTRATION DE LA VARIABLE DÉPENDANTE DU TEMPS À CHANGEMENT UNIQUE (*tableau 1*)

Configuration 1 (fig. 1). — Considérons le premier temps de survie, celui du sujet A, $t = 5,7$ mois. À ce temps, les cinq sujets A, B, C, D et E sont encore à risque de présenter l'événement ; les sujets A et C sont exposés puisque leur lipodystrophie survient avant 5,7 mois ; le sujet E n'est pas exposé puisque sa lipodystrophie survient après 5,7 mois : les sujets B et D ne sont pas exposés car ils ne déclarent pas de lipodystrophie au cours de leur suivi. Considérons ensuite le temps de survie du sujet C, $t = 17,9$ mois. À ce temps, les sujets C, D et E sont à risque ; les sujets C et E sont exposés (la lipodystrophie du sujet E survient avant 17,9 mois) ; le sujet D n'est pas exposé. Considérons enfin le temps de survie du sujet D, $t = 23,4$ mois. À ce temps, les sujets D et E sont à risque ; le sujet E est exposé ; le sujet D n'est pas exposé.

Configuration 2 (fig. 2). — Dans cette configuration, la lipodystrophie du sujet C survient *après* l'événement du sujet A. Par conséquent, au temps de survie du sujet A considéré ($t = 5,7$ mois), le sujet C n'est pas exposé. La considération des deux temps de survie suivants ($t = 17,9$ mois et $t = 23,4$ mois respectivement pour les sujets C et D) ne modifie pas les expositions prises en compte par rapport à la configuration 1.

Comparaison des deux configurations. — Quelle que soit la configuration, 1 ou 2, l'exposition

¹ Troubles de la répartition des graisses dans le corps.



L’abscisse correspond au délai écoulé (en mois) depuis la date d’entrée dans l’étude. Les traits horizontaux pleins (—) correspondent aux suivis des sujets A, C et D ayant présenté l’événement au terme de leur suivi (le trait vertical en pointillés permet de repérer les temps auxquels surviennent les événements) ; les traits horizontaux en pointillés (- -) correspondent aux suivis des sujets B et E censurés au terme de leur suivi. Les carrés (■) représentent les survenues de lipodystrophie pour les sujets A, C et E ; les losanges (◆) représentent les visites des sujets ; le chiffre sous chaque losange représente la valeur du taux de CD4 mesuré à la visite concernée (en nombre de CD4 par mm³). Seules les données du sujet C sont modifiées entre les figures 1 et 2 : délai de survenue de lipodystrophie, emplacement de la troisième visite, et valeur du taux de CD4 mesuré à cette nouvelle visite.

moyenne sur l’ensemble des temps considérés des sujets ayant présenté l’événement reste identique (0,67). Il existe, en revanche, une différence d’exposition moyenne chez les sujets à risque : 0,5 pour la configuration 1, et 0,4 pour la configuration 2. L’exposition moyenne des sujets à risque étant moins élevée dans la configuration 2 que dans la configuration 1, l’exposition est plus péjorative (le logiciel fournit les valeurs des paramètres β de 0,63 et 1,03 respectivement pour les configurations 1 et 2 ; cf. Annexe 3). Dans ces deux configurations, nous avons vu que le délai

entre la date d’origine du suivi et la date de survenue de l’exposition a été pris en compte. Cependant, ce ne sont pas les valeurs exactes de ce délai qui sont prises en compte, mais la position relative des dates de survenue d’exposition par rapport aux dates des événements. En effet, les résultats n’auraient pas été modifiés si, par exemple, le sujet E avait eu une lipodystrophie 6 mois après la date d’origine du suivi au lieu de 10,7 mois, puisque dans ces deux cas, la lipodystrophie du sujet E survenait entre l’événement du sujet A et celui du sujet C.

TABLEAU I. — Valeur de l'exposition prise en compte en tant que variable dépendante du temps à changement unique en fonction des temps d'événement considérés et des deux configurations.

Temps d'événement considérés	Exposition chez les sujets présentant l'événement		Exposition chez les sujets à risque au moment de l'événement	
	Config. 1	Config. 2	Config. 1	Config. 2
5,7 mois	1 (A)	1 (A)	1 (A), 0 (B), 1 (C), 0 (D), 0 (E)	1 (A), 0 (B), 0 (C), 0 (D), 0 (E)
17,9 mois	1 (C)	1 (C)	1 (C), 0 (D), 1 (E)	1 (C), 0 (D), 1 (E)
23,4 mois	0 (D)	0 (D)	0 (D), 1 (E)	0 (D), 1 (E)
Exposition moyenne	2/3 = 0,67	2/3 = 0,67	5/10 = 0,5	4/10 = 0,4

Le chiffre « 0 » ou « 1 » dans les colonnes « Config. 1 » et « Config. 2 » correspond à la valeur de l'exposition du sujet dont la lettre figure entre parenthèses, prise en compte dans le modèle de Cox. Ce chiffre vaut « 1 » si le sujet a été exposé (survenue d'une lipodystrophie) avant l'événement considéré, et « 0 » sinon. Sur la dernière ligne du tableau figure la moyenne de l'exposition sur l'ensemble des trois temps d'événement

TABLEAU II. — Valeur de l'exposition prise en compte en tant que variable dépendante du temps à changements multiples en fonction des temps d'événement considérés et des deux configurations.

Temps d'événement considérés	Exposition chez les sujets présentant l'événement		Exposition chez les sujets à risque au moment de l'événement	
	Config. 1	Config. 2	Config. 1	Config. 2
5,7 mois	1070 (A)	1070 (A)	1 070 (A), 703 (B), 1 435 (C), 312 (D), 732 (E)	1 070 (A), 703 (B), 1 267 (C), 312 (D), 732 (E)
17,9 mois	1206 (C)	1206 (C)	1 206 (C), 962 (D), 845 (E)	1 206 (C), 962 (D), 845 (E)
23,4 mois	962 (D)	962 (D)	962 (D), 1 237 (E)	962 (D), 1 237 (E)
Exposition moyenne	977	977	946	930

ILLUSTRATION DE LA VARIABLE DÉPENDANTE DU TEMPS À CHANGEMENTS MULTIPLES (tableau II)

Configuration 1 (fig. 1). — Considérons le premier temps de survie, celui du sujet A, $t = 5,7$ mois. À ce temps, la mesure du taux de CD4 des cinq sujets à risque A, B, C, D et E la plus proche précédant l'événement est respectivement 1 070, 703, 1 435, 312 et 732 CD4 par mm^3 . Considérons ensuite le temps de survie du sujet C, $t = 17,9$ mois. À ce temps, la mesure du taux de CD4 des trois sujets à risque C, D et E la plus proche précédant l'événement est respectivement 1 206, 962 et 845 CD4 par mm^3 . Considérons enfin le temps de survie du sujet D, $t = 23,4$ mois. À ce temps, la mesure du taux de CD4 des deux sujets à risque D et E la plus proche précédant l'événement est respectivement 962 et 1 237 CD4 par mm^3 .

Configuration 2 (fig. 2). — La modification entre les configurations 1 et 2 concerne la date de

la troisième visite du sujet C : celle-ci se situe non plus *avant* mais *après* l'événement du sujet A. Lorsque l'on considère le temps de survie du sujet A, $t = 5,7$ mois, la mesure du taux de CD4 du sujet C la plus proche précédant l'événement n'est plus 1 435 mais 1 267 CD4 par mm^3 . La considération des deux autres temps de survie ($t = 17,9$ mois et $t = 23,4$ mois respectivement pour les sujets C et D) ne modifie pas les expositions prises en compte par rapport à la configuration 1.

Comparaison des deux configurations. — L'exposition moyenne des sujets ayant présenté l'événement est identique dans les deux configurations (977 CD4 par mm^3). En revanche, celle concernant les sujets à risque est inférieure dans la configuration 2 (930 CD4 par mm^3) par rapport à la configuration 1 (946 CD4 par mm^3). Par conséquent, l'effet de l'exposition tend davantage

vers un effet péjoratif dans cette configuration 2 par rapport à la configuration 1 (le logiciel fournit les valeurs des paramètres β , pour une augmentation de 1 CD4, de respectivement 0,00154 et 0,00235 pour les configurations 1 et 2 ; cf. Annexe 4).

Nous pouvons remarquer dans cet exemple que certaines valeurs de l'exposition n'ont jamais participé au calcul du paramètre β . C'est le cas par exemple des premières valeurs mesurées pour les cinq sujets : leur deuxième mesure du taux de CD4 ayant lieu avant la date de survenue du premier événement, les premières mesures réalisées ne sont jamais prises en compte. En particulier, la présence d'une donnée manquante sur le taux de CD4 à la première visite du sujet A n'a pas, dans ce cas précis, posé de problème quant à l'estimation du paramètre.

LES DONNÉES MANQUANTES

Les problèmes posés par les données manquantes ont été largement décrits dans la littérature [19], et plus particulièrement dans le cas des variables dépendantes du temps [11, 20]. Little et Rubin (1987) ont montré que des biais importants peuvent se produire si le processus des données manquantes n'est pas aléatoire [21] ; on parle alors de données manquantes non ignorables [22].

Nous avons vu que les dates des mesures de l'exposition intervenaient pour « situer » l'exposition par rapport à la date de l'événement considéré. Une imprécision sur ces dates peut entraîner des biais d'information dans l'estimation des paramètres. Au moins deux cas de figure peuvent se présenter : a) les dates de mesure de l'exposition ne sont pas connues précisément, les dates attribuées sont par conséquent arbitrairement celles des visites ou du milieu entre deux visites ; b) les dates de mesure de l'exposition sont connues précisément, mais l'exposition est mesurée « trop longtemps » avant l'événement (il faut alors faire l'hypothèse d'une exposition constante entre la date de la mesure et la date de l'événement). Ces biais peuvent cependant être considérés *a priori* comme non différentiels si les visites sont standardisées et si les mesures sont réalisées en insu. Ainsi, dans l'exemple que nous avons choisi, nous avons considéré que la date de sur-

venue de la lipodystrophie était, pour simplifier, égale à la date de la visite à laquelle elle était diagnostiquée.

Dans l'exemple que nous avons utilisé, il n'y avait pas de données manquantes concernant la présence ou l'absence de lipodystrophie. Par conséquent, lorsqu'un sujet n'avait jamais présenté de signes de lipodystrophie au cours de son suivi, nous avons pu attribuer la valeur « . » pour la variable relative au délai entre la date d'origine et la survenue d'une lipodystrophie sans qu'il y ait d'ambiguïté. Dans le cas de données manquantes, il serait envisageable de créer une variable binaire *lipo_manquante* qui vaudrait « 1 » si la donnée sur la présence ou l'absence de lipodystrophie manque, et « 0 » si cette donnée est présente, puis rajouter dans la programmation l'instruction « AND lipo_manquante = 0 » à la ligne 3 du programme situé en annexe 3. Si la donnée manque de manière non aléatoire [21, 22], l'exclusion des sujets dont la donnée manque ou la création d'une modalité correspondant à ces sujets pose des problèmes dans l'estimation du paramètre associé à l'exposition (biais d'estimation) [19, 23]. Des méthodes plus sophistiquées mais meilleures en termes d'estimation et de variance, telles que les imputations multiples, ont été bien décrites dans la littérature statistique [24, 25, 26]. Ces méthodes ne sont, en revanche, pas encore d'un usage courant avec les logiciels utilisés en épidémiologie.

Dans le cas des variables à changements multiples, si l'on note t_k le délai entre la date d'origine de l'étude et la $k^{\text{ème}}$ mesure effectuée à la $k^{\text{ème}}$ visite, l'hypothèse formulée dans l'exemple précédent considèrerait que l'exposition était constante dans l'intervalle $[t_k ; t_{k+1}[$. Pour certains marqueurs biologiques et selon l'espacement des visites du protocole, cette hypothèse peut être considérée comme forte. Si de plus l'exposition n'est pas mesurée à la $k^{\text{ème}}$ visite prévue par le protocole (le sujet n'est pas venu pour le prélèvement), l'intervalle sur lequel l'exposition est considérée comme constante devient $[t_{k-1} ; t_{k+1}[$. Une option simple à programmer consiste à estimer l'exposition manquante par interpolation entre deux mesures. Une solution préférable suggérée dans la littérature [27, 28] consiste à estimer les données manquantes par imputations simples ou

multiples à l'aide d'une modélisation à effets mixtes [29].

Notons enfin que SAS gère les données manquantes sur l'exposition traitée en tant que variable dépendante du temps à changements multiples de manière différente selon la position dans le temps des données manquantes. Supposons un sujet i présentant un événement au temps t_j . Si, pour ce sujet, l'exposition est manquante au temps pertinent (dans notre exemple, la visite précédent l'événement), la vraisemblance V_j associée à ce temps n'est pas calculée dans la mesure où le terme « $h_i(t_j)$ » placé au numérateur est manquant. Le logiciel considère alors que le sujet i n'a pas présenté l'événement, et laisse ce sujet dans le groupe des sujets censurés.

LIMITES ET SOLUTIONS ALTERNATIVES

Les précautions à prendre lors de l'introduction d'une variable dépendante du temps en analyse de survie ont déjà été soulignées [16]. Il faut notamment s'attacher, comme dans toute analyse en épidémiologie, à expliciter le statut de la variable dépendante du temps, facteur de risque ou facteur intermédiaire dans la chaîne causale. Dans le cas de sujets infectés par le VIH suivis dans une cohorte, on pourrait étudier l'effet de la mise sous monothérapie sur le risque de survenue d'un sida, exposition traitée en variable dépendante du temps. Le résultat du modèle peut montrer un effet péjoratif du traitement, dans la mesure où la mise sous traitement est un marqueur de la dégradation de l'état de santé du sujet. L'ajustement sur certaines caractéristiques de l'état de santé du sujet permettrait sûrement de trouver des résultats allant dans le sens opposé à celui trouvé initialement. Cependant, Robins (1989) a montré que l'estimation de l'effet de la mise sous traitement antirétroviral peut être biaisée si ces caractéristiques sont liées à la mise sous traitement et liées à la maladie (par exemple, le taux de CD4 recueilli au cours du suivi) [30]. C'est dans cet objectif de prise en compte de tels facteurs que les modèles structuraux marginaux ont été développés [31, 32], en particulier dans l'étude de l'effet du traitement antirétroviral sur la survenue d'un sida ou d'un décès chez des patients infectés par le VIH [33, 34].

D'autres modèles d'analyse de survie basés sur le modèle de Cox sont capables de faire intervenir des modifications de l'exposition au cours du suivi du sujet, tels que le modèle « Cox-Poisson » [35] dans le cas où la date de l'événement est précisément connue, ou le modèle « Cox-discret » [36] dans le cas où les temps de survie sont intrinsèquement discrets ou bien regroupés en intervalles de temps discrets (par exemple, la probabilité de grossesse sur un temps de survenue exprimé en nombre de cycles [37]). Enfin, des extensions du modèle de Cox ont été décrites. Hastie et Tibshirani (1990) ont fait intervenir des méthodes de lissage pour des expositions quantitatives, en introduisant un vecteur de fonctions $f(X)$ (fonctions splines cubiques) à la place du vecteur βX dans l'expression du modèle de Cox : $h(t) = \lambda_0(t) \exp\{f(X)\}$, où X est un vecteur de variables quantitatives et β le vecteur des paramètres associés [38]. Cette méthode exploratoire, qui permet de ne plus faire l'hypothèse d'une relation linéaire entre l'exposition quantitative et le paramètre β , s'adapte très bien au cas des variables quantitatives dépendantes du temps à changements multiples. Kooperberg et Clarkson (1997) ont quant à eux amélioré un modèle de régression de force de mortalité (« hazard regression » (HARE)). Ce modèle HARE a été développé initialement par Kooperberg *et al.* [39] et comporte lui aussi des fonctions cubiques splines. L'amélioration de ce modèle permet désormais de prendre en compte des variables dépendantes du temps à changements multiples (l'exemple par ailleurs cité dans l'article est celui du taux de CD4 recueilli au cours du suivi) [40].

Récemment, une nouvelle catégorie de modèle (les modèles joints) ont été développés pour prendre en compte l'évolution d'un marqueur au cours du suivi d'un sujet dans l'étude de la survenue d'un événement en analyse de survie [41-43]. Ces modèles prennent en compte, de façon conjointe, l'évolution d'un marqueur mesuré plusieurs fois pour un même sujet (modélisation longitudinale à effets mixtes), et un processus de survenue d'événement (modélisation de Cox). Guo et Carlin (2004) présentent ce type de modélisation sur des données réelles en utilisant les logiciels SAS et WinBUGS (logiciel disponible gratuitement sur le site Internet <http://www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml>). Les auteurs

ont rendu disponible leur programmation, réalisée avec les deux logiciels, sur le site Internet <http://www.biostat.umn.edu/~brad/software.html> [44]. Cette disponibilité des programmes va rendre désormais la modélisation jointe beaucoup plus accessible.

Pour conclure, l'introduction d'une variable dépendante du temps dans un modèle de Cox permet de prendre en compte une modification de l'exposition initiale au cours du suivi d'un sujet, ce que ne permet pas l'introduction en tant que variable « fixe ». Il est cependant nécessaire de reconnaître son statut (par exemple de distinguer entre facteur de risque et facteur intermédiaire) pour éviter une mauvaise interprétation des résultats. L'utilisation des diagrammes de causalité peut fournir une aide précieuse dans cette réflexion [45, 46]. Une fois cette réflexion réalisée, l'introduction d'une exposition en tant que variable dépendante du temps permet malgré tout de répondre à la majorité des questions posées lorsque les expositions varient au cours du temps.

REMERCIEMENTS : Nous tenons particulièrement à remercier Jean-Baptiste HUBERT pour son aide très précieuse dans ce domaine, et le groupe méthodologie des unités INSERM-INED U569 et U149 pour nous avoir encouragés dans la rédaction de cet article (membres du groupe : Pierre-Yves ANCEL, Farouy BOUFASSA, Jean BOUYER, Loïc DESQUILBET, Béatrice DUCOT, Hélène GOULARD, Babak KHOSHNOOD, Elise DE LA ROCHEBROCHARD, Jérôme LECHENADEC, Yoann MADEC, Laurence MEYER, Caroline MOREAU, Audrey MULLER, Rémy SLAMA, Jean-Paul TÉGLAS, Josiane WARSAWSKI, Jenifer ZEITLIN).

RÉFÉRENCES

1. Cox DR. Regression models and life tables (with discussion). *J R Statist Soc B* 1972; 63: 269-76.
2. O'Quigley J, Xu R. Goodness-of-fit in survival analysis. New York, NY: John Wiley, 1998.
3. Quantin C, Asselain B, Moreau T. Le modèle de Cox : limites et extensions. *Rev Épidemiol Santé Publique* 1990 ; 38 : 341-56.
4. Contal C, Mallet A. Mise à l'épreuve pratique du modèle de Cox. *Rev Épidemiol Santé Publique* 2000 ; 48 : 490-501.
5. Clark DA, Stinson EB, Griep RB, Schroeder JS, Shumway NE, Harrison DC. Cardiac transplantation in man. VI. Prognosis of patients selected for cardiac transplantation. *Ann Intern Med* 1971; 75: 15-21.
6. Gail MH. Does cardiac transplantation prolong life? A reassessment. *Ann Intern Med* 1972; 76: 815-7.
7. Crowley J, Hu M. Covariance analysis of heart transplant survival data. *Journal of the American Statistical Association* 1977; 72: 27-36.
8. Dvorchik I, Subotin M, Demetris AJ, *et al.* Effect of liver transplantation on inflammatory bowel disease in patients with primary sclerosing cholangitis. *Hepatology* 2002; 35: 380-4.
9. Feldman JG, Gange SJ, Bacchetti P, *et al.* Serum albumin is a powerful predictor of survival among HIV-1-infected women. *J Acquir Immune Defic Syndr* 2003; 33: 66-73.
10. Taffé P, Rickenbach M, Hirschel B, *et al.* Impact of occasional short interruptions of HAART on the progression of HIV infection: results from a cohort study. *Aids* 2002; 16: 747-55.
11. Altman DG, De Stavola BL. Practical problems in fitting a proportional hazards model to data with updated measurements of the covariates. *Stat Med* 1994; 13: 301-41.
12. Allison PD. *Survival analysis using the SAS® System: a practical guide.* Cary, NC: SAS Institute Inc, 1995.
13. Hosmer DWJ, Lemeshow S. *Applied survival analysis. Regression modeling of time to event data.* New York, NY: John Wiley & Sons, 1999.
14. Breslow N. Covariance analysis of censored survival data. *Biometrics* 1974; 30: 89-99.
15. Efron B. The efficiency of Cox's likelihood function for censored data. *Journal of the American Statistical Association* 1977; 72: 557-65.
16. Kalbfleisch JD, Prentice RL. *The statistical analysis of failure time data.* New York: Wiley, 1980.
17. Andersen PK. Repeated assessment of risk factors in survival analysis. *Stat Methods Med Res* 1992; 1: 297-315.
18. Desquilbet L, Deveau C, Goujard C, Hubert JB, Derouineau J, Meyer L. Increase in at-risk sexual behaviour among HIV-1-infected patients followed in the French PRIMO cohort. *Aids* 2002; 16: 2329-33.
19. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. *Am J Epidemiol* 1995; 142: 1255-64.
20. Halabi S, Wun CC, Davis BR. Analysis of survival data with missing measurements of a time-dependent binary covariate. *J Biopharm Stat* 2003; 13: 253-70.
21. Little RJA, Rubin DB. *Statistical analysis with missing data.* New York: John Wiley & Sons, 1987.
22. Jacqmin-Gadda H, Commenges D, Dartigues J-F. Analyse de données longitudinales gaussiennes comportant des données manquantes sur la variable à expliquer. *Rev Épidemiol Santé Publique* 1999; 47: 525-34.

23. Vach W, Blettner M. Biased estimation of the odds ratio in case-control studies due to the use of ad hoc methods of correcting for missing values for confounding variables. *Am J Epidemiol* 1991; 134: 895-907.
24. Faris PD, Ghali WA, Brant R, Norris CM, Galbraith PD, Knudtson ML. Multiple imputation versus data enhancement for dealing with missing data in observational health care outcome analyses. *J Clin Epidemiol* 2002; 55: 184-91.
25. Van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med* 1999; 18: 681-94.
26. Molenberghs G, Williams PL, Lipsitz SR. Prediction of survival and opportunistic infections in HIV-infected patients: a comparison of imputation methods of incomplete CD4 counts. *Stat Med* 2002; 21: 1387-408.
27. Wu H, Wu L. A multiple imputation method for missing covariates in non-linear mixed-effects models with application to HIV dynamics. *Stat Med* 2001; 20: 1755-69.
28. Tsiatis AA, DeGruttola V, Wulfsohn MS. Modeling the relationship of survival to longitudinal data with error: applications to survival and CD4 counts in patients with AIDS. *J Am Stat Assoc* 1995; 90: 27-37.
29. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982; 38: 963-74.
30. Robins J. The control of confounding by intermediate variables. *Stat Med* 1989; 8: 679-701.
31. Brumback BA, Hernan MA, Haneuse SJ, Robins JM. Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. *Stat Med* 2004; 23: 749-67.
32. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000; 11: 550-60.
33. Cole SR, Hernan MA, Robins JM, *et al.* Effect of highly active antiretroviral therapy on time to acquired immunodeficiency syndrome or death using marginal structural models. *Am J Epidemiol* 2003; 158: 687-94.
34. Hernan MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* 2000; 11: 561-70.
35. Clayton D. The analysis of event history data: a review of progress and outstanding problems. *Stat Med* 1988; 7: 819-41.
36. Prentice RL, Gloeckler LA. Regression analysis of grouped survival data with application to breast cancer data. *Biometrics* 1978; 34: 57-67.
37. Bonde JP, Ernst E, Jensen TK, *et al.* Relation between semen quality and fertility: a population-based study of 430 first-pregnancy planners. *Lancet* 1998; 352: 1172-7.
38. Hastie T, Tibshirani R. Exploring the nature of covariate effects in the proportional hazards model. *Biometrics* 1990; 46: 1005-16.
39. Kooperberg C, Stone CJ, Truong YK. Hazard regression. *J Am Stat Assoc* 1995; 90: 78-94.
40. Kooperberg C, Clarkson DB. Hazard regression with interval-censored data. *Biometrics* 1997; 53: 1485-94.
41. Song X, Davidian M, Tsiatis AA. A semiparametric likelihood approach to joint modeling of longitudinal and time-to-event data. *Biometrics* 2002; 58: 742-53.
42. Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics* 1997; 53: 330-9.
43. Dupuy JF, Mesbah M. Joint modeling of event time and nonignorable missing longitudinal data. *Lifetime Data Anal* 2002; 8: 99-115.
44. Guo X, Carlin P. Separate and joint modeling of longitudinal and event time data using standard computer packages. *Am Stat* 2004; 58: 1-9.
45. Cole SR, Hernan MA. Fallibility in estimating direct effects. *Int J Epidemiol* 2002; 31: 163-5.
46. Hernan MA, Hernandez-Diaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *Am J Epidemiol* 2002; 155: 176-84.

ANNEXE 1. — La procédure PROC PHREG dans SAS

```

PROC PHREG DATA = fic_donnees ;
MODEL survie*even (0) = exposition1 exposition2... ;
[...]
RUN ;

```

La procédure SAS pour analyser des données de survie avec le modèle de Cox est la procédure PROC PHREG. Le fichier de données (*fic_donnees*) doit contenir une ligne par sujet et au moins les deux variables suivantes : a) la variable *survie* valant le délai écoulé entre la date d'origine du suivi et la date de survenue de l'événement ou de la censure, et b) la variable *even* valant « 1 » si le sujet a présenté l'événement, et « 0 » sinon. Les variables *exposition1*, *exposition2*, ... sont des variables incluses dans le modèle multivarié, qui peuvent être soit qualitatives, soit quantitatives. Le symbole « [...] » correspond aux instructions optionnelles ou supplémentaires de la procédure. Les instructions optionnelles sont, par exemple, celles qui permettent d'analyser une sous-population de l'échantillon (grâce à l'instruction « WHERE »), ou bien celles permettant d'explicitier le processus interne de SAS définissant les variables dépendantes du temps. Les instructions supplémentaires sont, par exemple, celles qui définissent une variable en tant que variable dépendante du temps.

ANNEXE 2. — Fichiers de données

a) *Présentation des fichiers de données*

Le fichier de données initial comporte 5 sujets (sujets A à E) issus de la cohorte PRIMO. Le nombre de visites par sujet au cours du suivi varie de 3 à 7. La structure du fichier de données nécessaire pour prendre en compte une exposition en tant que variable dépendante du temps pour le logiciel SAS est celle présentée dans la suite de cette annexe : une ligne par sujet, les variables de suivi étant placées en colonnes.

Tous les délais sont exprimés en mois. La variable *id* est la variable relative à l'identifiant ; la variable *even* vaut « 1 » si le sujet a déclaré l'événement (une relation sexuelle sans préservatif), et « 0 » s'il ne l'a pas déclarée ; la variable *survie* vaut le délai écoulé entre la date d'origine de l'étude et l'événement ou la censure ; la variable *delai_lipo* vaut le délai écoulé entre la date d'origine de l'étude et la date de survenue d'une lipodystrophie, ou bien vaut « . » si le sujet n'a pas présenté de lipodystrophie avant l'événement ou la censure ; les variables *suivi_i* ($i \in [1 ; 7]$) valent le délai écoulé entre la date d'origine de l'étude et la date de la visite n°*i* ; les variables *CD4_i* ($i \in [1 ; 7]$) valent le taux de CD4 mesuré à la visite n°*i* (en CD4 par mm³). Si un sujet n'est plus suivi au delà de la *k*^{ème} mesure, les valeurs des variables *suivi_(k+1)* jusqu'à *suivi_7*, de même que celles des variables *CD4_(k+1)* jusqu'à *CD4_7* seront manquantes.

Entre les configurations 1 et 2, le fichier de données initial a été modifié seulement pour le sujet C : la lipodystrophie survient au bout de 7,4 mois (au lieu de 3,4 mois pour la configuration 1), et la visite n° 3 a lieu 8,2 mois après la date d'entrée dans l'étude (au lieu de 5,2 mois) avec une valeur de CD4 de 1 338 CD4 par mm³ mesuré à cette nouvelle visite (au lieu de 1 435 CD4 par mm³).

b) Fichier de données initial, configuration 1

id	even	survie	delai_lipo	suiwi_1	suiwi_2	suiwi_3	suiwi_4	suiwi_5	suiwi_6	suiwi_7	CD4_1	CD4_2	CD4_3	CD4_4	CD4_5	CD4_6	CD4_7
A	1	5.7	4.1	0.8	2.6	5.7	1 070	886
B	0	12.4	.	0.9	3.0	6.5	12.4	.	.	.	664	703	834	440	.	.	.
C	1	17.9	7.4	1.1	2.1	8.2	12.3	17.9	.	.	792	1 267	1 338	1 206	1 386	.	.
D	1	23.4	.	1	2.8	6.4	11.5	17.2	23.4	.	560	312	882	1 218	962	660	.
E	0	29.4	10.7	0.9	3.2	7.0	13.2	18.8	25.5	29.4	654	732	645	845	1 237	1 390	961

c) Fichier de données modifié, configuration 2

Les modifications entre les deux fichiers de données concernent le sujet C ; elles apparaissent en gras dans le tableau ci-dessous.

id	even	survie	delai_lipo	suiwi_1	suiwi_2	suiwi_3	suiwi_4	suiwi_5	suiwi_6	suiwi_7	CD4_1	CD4_2	CD4_3	CD4_4	CD4_5	CD4_6	CD4_7
A	1	5.7	4.1	0.8	2.6	5.7	1 070	886
B	0	12.4	.	0.9	3.0	6.5	12.4	.	.	.	664	703	834	440	.	.	.
C	1	17.9	3.4	1.1	2.1	5.2	12.3	17.9	.	.	792	1 267	1 435	1 206	1 386	.	.
D	1	23.4	.	1	2.8	6.4	11.5	17.2	23.4	.	560	312	882	1 218	962	660	.
E	0	29.4	10.7	0.9	3.2	7.0	13.2	18.8	25.5	29.4	654	732	645	845	1 237	1 390	961

ANNEXE 3. — Programmation dans SAS d'une variable dépendante du temps à changement unique, et résultats fournis par le logiciel

a) Programmation et commentaires

```

PROC PHREG DATA = config_1 ;/* ou config_2 */
MODEL survie*even(0) = lipo_CU ;
IF delai_lipo < survie AND delai_lipo ne.
THEN lipo_CU = 1 ; ELSE lipo_CU = 0 ;
FILE log ;
PUT 'Sujet = ' id ; PUT 'survie considérée = ' survie ;
PUT 'délai de lipodystrophie du sujet = ' delai_lipo ;
PUT 'Valeur de l'exposition prise en compte pour ce sujet = ' lipo_CU ;
RUN ;

```

La variable dépendante du temps à changement unique (notée *lipo_CU* dans la programmation) se définit obligatoirement au sein de la procédure PROC PHREG. Cette définition, située à la ligne 4, n'est exécutée que chez les sujets à risque au temps d'événement considéré. De plus, à la ligne 3, la valeur que prend la variable *survie* n'est pas celle de chaque sujet, mais celle du temps de survie de l'événement considéré (cf « contenu de la fenêtre 'Journal' de SAS » dans la suite de cette annexe). Si le nombre d'événements est k , la variable *survie* prendra k valeurs. Un sujet qui a été exposé au cours de son suivi est considéré comme exposé (*lipo_CU* = « 1 » dans le modèle) lorsque la valeur de *delai_lipo* pour ce sujet est strictement inférieure au temps de survie considéré. Le sujet est considéré comme non exposé au moment de l'événement considéré si la valeur de *delai_lipo* de ce sujet est supérieure ou égale au temps d'événement considéré (la lipodystrophie du sujet est apparue au moment ou après l'événement considéré), ou si le sujet n'a jamais présenté de lipodystrophie (on fait alors l'hypothèse qu'il n'y a pas de données manquantes sur la présence ou l'absence de lipodystrophie). Les instructions (optionnelles) qui figurent aux lignes 5 à 8 permettent de faire apparaître le processus d'affectation de la valeur « 0 » ou « 1 » de l'exposition prise en compte dans les analyses pour chaque sujet, à chaque temps d'événement considéré. L'instruction « FILE LOG ; » à la ligne 5 définit la localisation de ce processus, ici dans la fenêtre « Journal » de SAS.

b) Contenu de la fenêtre « Journal » de SAS

Pour chacune des 2 configurations, nous avons sélectionné la considération du temps d'événement du sujet A : 5,7 mois (la considération pour laquelle il existe une différence de prise en compte de l'exposition entre les deux configurations, signalée en gras ci-dessous).

Configuration 1

[...]

Sujet = A

survie considérée = 5.7

délai de lipodystrophie du sujet = 4.1

Valeur de l'exposition prise en compte pour ce sujet = 1

Sujet = B

survie considérée = 5.7

délai de lipodystrophie du sujet = .

Valeur de l'exposition prise en compte pour ce sujet = 0

Sujet = C

survie considérée = 5.7

délai de lipodystrophie du sujet = 3.4

Valeur de l'exposition prise en compte pour ce sujet = 1

Sujet = D

survie considérée = 5.7

délai de lipodystrophie du sujet = .

Valeur de l'exposition prise en compte pour ce sujet = 0

Sujet = E

survie considérée = 5.7

délai de lipodystrophie du sujet = 10.7

Valeur de l'exposition prise en compte pour ce sujet = 0

Configuration 2

[...]

Sujet = A

survie considérée = 5.7

délai de lipodystrophie du sujet = 4.1

Valeur de l'exposition prise en compte pour ce sujet = 1

Sujet = B

survie considérée = 5.7

délai de lipodystrophie du sujet = .

Valeur de l'exposition prise en compte pour ce sujet = 0

Sujet = C**survie considérée = 5.7****délai de lipodystrophie du sujet = 7.4**

Valeur de l'exposition prise en compte pour ce sujet = 0

Sujet = D

survie considérée = 5.7

délai de lipodystrophie du sujet = .

Valeur de l'exposition prise en compte pour ce sujet = 0

Sujet = E

survie considérée = 5.7

délai de lipodystrophie du sujet = 10.7

Valeur de l'exposition prise en compte pour ce sujet = 0

***c) Paramètres et RR associés à la lipodystrophie
dans le modèle de Cox
(fenêtre « Sortie » de SAS)***Configuration 1

Analysis of maximum likelihood estimates						
Variable	DDL	Parameter estimate	Standard error	Chi-Square	Pr > ChiSq	Hazard ratio
lipo_CU	1	0.63289	1.25085	0.256	0.6129	1.883

Configuration 2

Analysis of maximum likelihood estimates						
Variable	DDL	Parameter estimate	Standard error	Chi-Square	Pr > ChiSq	Hazard ratio
lipo_CU	1	1.03464	1.33158	0.6037	0.4372	2.814

ANNEXE 4. — Programmation dans SAS d'une variable dépendante du temps à changements multiples, et résultats fournis par le logiciel

a) Programmation et commentaires

```

PROC PHREG DATA = config_1 ; /* ou config_2 */
MODEL survie*even(0) = CD4_CM ;
ARRAY suivi_col[7] suivi_1-suivi_7 ;
ARRAY CD4_col[7] CD4_1-CD4_7 ;
DO i = 1 TO 7 ;
IF suivi_col[i] < survie AND suivi_col[i] ne.
THEN CD4_CM = CD4_col[i] ;
FILE log ;
PUT 'Sujet = ' id ; PUT 'survie considérée = ' survie ;
PUT 'Numéro de la mesure = ' i ;
PUT 'Délai de mesure des CD4 = ' suivi_col[i] ;
PUT 'Valeur des CD4 à cette mesure = ' CD4_col[i] ;
PUT 'Valeur de l'exposition prise en compte pour ce sujet = ' CD4_CM ;
END ;
RUN ;

```

De la même manière que la variable dépendante du temps à changement unique, la variable à changements multiples se définit uniquement au sein de la procédure PROC PHREG. La définition de cette variable (notée $CD4_CM$ dans la programmation) se trouve à la ligne 7 du programme. La boucle située aux lignes 5 à 14 s'exécute seulement pour les sujets à risque au temps d'événement considéré. Cette boucle permet de sélectionner la dernière mesure telle que la condition située à la ligne 6 est encore vérifiée : tant que le délai entre la date d'origine du suivi et la $i^{\text{ème}}$ mesure est strictement inférieur au temps d'événement considéré, la variable $CD4_CM$ prend la valeur de l'exposition de cette $i^{\text{ème}}$ mesure (cf. *Annexe 4*). Ainsi, la valeur de l'exposition d'un sujet utilisée pour le calcul du paramètre β sera celle mesurée la plus proche précédant l'événement considéré. Les instructions (optionnelles) qui figurent aux lignes 8 à 13 permettent de faire apparaître dans la fenêtre « Journal » de SAS le processus de sélection de la valeur de l'exposition prise en compte dans les analyses pour chaque sujet, à chaque temps d'événement considéré. Cette valeur est celle atteinte au terme de la boucle, ici pour $i = 7$.

b) Contenu de la fenêtre « Journal » de SAS

Nous avons sélectionné le processus décrit ci-dessus pour le sujet C, au temps d'événement considéré du sujet A (5,7 mois). La valeur de l'exposition du sujet C prise en compte dans les analyses pour le temps d'événement considéré de 5,7 mois est celle en face de « Valeur de l'exposition prise en compte pour ce sujet = ». Dans la configuration 1, l'exposition vaut 1 435 CD4 par mm^3 ; dans la configuration 2, elle vaut 1 267 CD4 par mm^3 .

Configuration 1

[...]

Sujet = C

survie considérée = 5.7

Numéro de la mesure (i) = 1

Délai de mesure des CD4 = 1.1

Valeur des CD4 à cette mesure = 792

Valeur de l'exposition prise en compte pour ce sujet = 792

Sujet = C

survie considérée = 5.7

Numéro de la mesure (i) = 2

Délai de mesure des CD4 = 2.1

Valeur des CD4 à cette mesure = 1267

Valeur de l'exposition prise en compte pour ce sujet = 1267

Sujet = C

survie considérée = 5.7

Numéro de la mesure (i) = 3

Délai de mesure des CD4 = 5.2

Valeur des CD4 à cette mesure = 1435

Valeur de l'exposition prise en compte pour ce sujet = 1435

Sujet = C

survie considérée = 5.7

Numéro de la mesure (i) = 4

Délai de mesure des CD4 = 12.3

Valeur des CD4 à cette mesure = 1206

Valeur de l'exposition prise en compte pour ce sujet = 1435

Sujet = C

survie considérée = 5.7

Numéro de la mesure (i) = 5

Délai de mesure des CD4 = 17.9

Valeur des CD4 à cette mesure = 1386

Valeur de l'exposition prise en compte pour ce sujet = 1435

Sujet = C

survie considérée = 5.7

Numéro de la mesure (i) = 6

Délai de mesure des CD4 = .

Valeur des CD4 à cette mesure = .

Valeur de l'exposition prise en compte pour ce sujet = 1435

Sujet = C

survie considérée = 5.7

Numéro de la mesure (i) = 7

Délai de mesure des CD4 = .

Valeur des CD4 à cette mesure = .

Valeur de l'exposition prise en compte pour ce sujet = 1435

[...]

Configuration 2

[...]

Sujet = C

survie considérée = 5.7

Numéro de la mesure (i) = 1

Délai de mesure des CD4 = 1.1

Valeur des CD4 à cette mesure = 792

Valeur de l'exposition prise en compte pour ce sujet = 792

Sujet = C

survie considérée = 5.7

Numéro de la mesure (i) = 2

Délai de mesure des CD4 = 2.1

Valeur des CD4 à cette mesure = 1267

Valeur de l'exposition prise en compte pour ce sujet = 1267

Sujet = C

survie considérée = 5.7

Numéro de la mesure (i) = 3

Délai de mesure des CD4 = 8.2

Valeur des CD4 à cette mesure = 1338

Valeur de l'exposition prise en compte pour ce sujet = 1267

Sujet = C

survie considérée = 5.7

Numéro de la mesure (i) = 4

Délai de mesure des CD4 = 12.3

Valeur des CD4 à cette mesure = 1206

Valeur de l'exposition prise en compte pour ce sujet = 1267

Sujet = C

survie considérée = 5.7

Numéro de la mesure (i) = 5

Délai de mesure des CD4 = 17.9

Valeur des CD4 à cette mesure = 1386

Valeur de l'exposition prise en compte pour ce sujet = 1267

Sujet = C

survie considérée = 5.7

Numéro de la mesure (i) = 6

Délai de mesure des CD4 = .

Valeur des CD4 à cette mesure = .

Valeur de l'exposition prise en compte pour ce sujet = 1267

Sujet = C

survie considérée = 5.7

Numéro de la mesure (i) = 7

Délai de mesure des CD4 = .

Valeur des CD4 à cette mesure = .

Valeur de l'exposition prise en compte pour ce sujet = 1267

[...]

*c) Paramètres et RR associés au taux de CD4 dans le modèle de Cox (fenêtre « Sortie » de SAS)*Configuration 1

Analysis of maximum likelihood estimates

Variable	DDL	Parameter estimate	Standard error	Chi-Square	Pr > ChiSq	Hazard ratio
CD4_CM	1	0.00154	0.0024	0.4134	0.5202	1.002

Configuration 2

Analysis of maximum likelihood estimates

Variable	DDL	Parameter estimate	Standard error	Chi-Square	Pr > ChiSq	Hazard ratio
CD4_CM	1	0.00235	0.00298	0.6225	0.4301	1.002
