

Initiation aux enquêtes quantitatives

Réunion du 22 juin 2005, Bruno Spire

Date de mise en ligne : 18 juillet 2005

I – Les enquêtes

Les enquêtes s'intéressent à l'« observationnel » par opposition aux essais cliniques qu'on dit « interventionnels ». Dans les enquêtes, on regarde « la vraie vie » sans intervenir. De manière générale, une enquête est plus simple à organiser qu'un essai clinique, mais ses résultats sont souvent plus compliqués à analyser.

Les différents types d'enquête :

1) Les enquêtes transversales

Elles correspondent à « une photo d'une situation à un moment donné ». Exemple : l'enquête Vespa est une enquête transversale. Elle regarde la situation de plusieurs centaines de séropositifs à un moment donné de leur vie. Elle permet d'obtenir une « photographie », à un temps « t », de la population des séropositifs en France.

Les avantages de l'enquête transversale : elle est simple à réaliser, on « enquête » une seule fois.

Des inconvénients de l'enquête transversale :

- Quelle est la représentativité de l'échantillon enquêté ? Qui enquête-t-on vraiment ? Qui a répondu à l'enquête ? Ou, au contraire, qui n'a pas répondu à l'enquête ? Il faut se poser la question de l'échantillon que l'on souhaite enquêter avant de mener l'enquête. Les répondants sont-ils représentatifs de la population que l'on souhaite étudier ? Lors de l'analyse des résultats, il faut veiller à interpréter les données en fonction de l'échantillon enquêté. Si l'échantillon ne correspond qu'à un groupe de personnes, on ne peut pas tirer des conclusions pour l'ensemble d'une population. Un biais classique des enquêtes transversales : le biais de fréquentation. Dans une enquête transversale à l'hôpital, on a plus de chance de tomber sur des usagers fréquents (sur-représentation des personnes qui n'ont pas trop de problèmes de prise en charge globale par opposition aux personnes qui ont plus de difficultés dans l'accès aux soins). Les enquêteurs sont alors contraints de « redresser » les résultats : on demande aux personnes interrogées à quelle fréquence elles viennent à l'hôpital, et on pondère les résultats par la fréquence des visites, i.e. on donne plus d'importance à celles qui ne viennent pas souvent et moins d'importance à celles qui viennent très souvent dans l'analyse.

- Des problèmes supplémentaires liés à la forme « questionnaire » peuvent survenir : écueils de langue, inquiétudes quant à la confidentialité, influence du conformisme social, auto-questionnaire versus questionnaire en face à face (on ne

répond pas de la même façon), formulation des questions... Exemple : sur des questions de prévention dans les rapports sexuels, on n'obtient pas les mêmes réponses pour un questionnaire dirigé avec des questions intrusives qu'avec un auto-questionnaire garantissant la confidentialité... La mise au point d'un questionnaire est un exercice complexe qui nécessite des compétences et des regards multiples pour éviter les biais et trouver les questions les plus adaptées aux informations que l'on cherche à recueillir.

- La seconde grande faiblesse des enquêtes transversales réside dans la (non) fiabilité des données rétrospectives. On ne peut pas déduire une évolution dans le temps de la comparaison des résultats de deux enquêtes transversales. Exemple : dans l'enquête A, M. Dupont est marié. Dans l'enquête B, M. Dupont est aussi marié. On ne peut pas déduire de ces deux enquêtes que M. Dupont est resté marié pendant les 10 ans qui séparent les deux enquêtes. M. Dupont peut avoir divorcé d'une première femme et en avoir épousé une autre après 8 ans de célibat. En somme, ces enquêtes ne permettent pas d'expliquer les évolutions dans le temps d'une situation donnée (les analyses de ces évolutions sont pourtant tentantes).

- La troisième grande limite des enquêtes transversales est relative à l'absence d'homogénéité de la durée d'exposition à quelque chose. Exemple : si on fait une enquête à un instant « t » sur le vécu de personnes prenant du Sustiva, on peut avoir dans l'échantillon des personnes qui prennent Sustiva depuis 3 mois et d'autres depuis 5 ans. Cette hétérogénéité doit être prise en compte (on peut poser la question dans le questionnaire : « depuis combien de temps prenez-vous Sustiva ? ») et constitue une caractéristique des enquêtes transversales par rapport aux cohortes (dans une cohorte, on a une homogénéité de la durée de traitement sous Sustiva, si le point d'entrée dans la cohorte est justement la mise sous Sustiva). Dans le cas de l'enquête Sustiva réalisée par le TRT-5, cette limite de l'enquête n'a pas permis, par exemple, d'évaluer une relation entre la prise de Sustiva et la dépression (car certains malades très dépressifs avaient déjà arrêté Sustiva avant d'être enquêtés ; d'autres l'avaient arrêté juste après l'enquête...).

De manière générale, les effets indésirables doivent préférentiellement être étudiés dans le cadre d'études longitudinales. On peut bien sûr tenter de récupérer quelques informations rétrospectivement, mais cela n'est pas idéal.

2) Les enquêtes longitudinales

Il s'agit d'enquêtes prospectives, qui vont être menées sur une période donnée. Dans les enquêtes longitudinales, on étudie une problématique commune à un groupe de personnes qui seront suivies sur plusieurs mois ou années. La principale difficulté de ce type d'enquête est représentée par « l'attrition » : au cours de l'enquête, on va perdre de vue des répondants pour des raisons multiples. Pour pallier ces pertes, les enquêteurs sont contraints de prévoir de grands échantillons à l'initiation. Ces enquêtes sont proches des cohortes. La différence entre les cohortes

et les enquêtes longitudinales sont, selon les définitions de l'ANRS :

- la cohorte dure un minimum de 3 ans ;
- dans la cohorte, au temps 0, il y a un événement commun aux personnes incluses (mise sous traitement, entrée pour la première fois à l'hôpital, etc.), ce qui n'est pas le cas dans les enquêtes longitudinales.

Exemple d'enquête longitudinale : l'enquête Oskar que lance Abbott pour observer la « perception patient » sous Kaletra pendant 9 mois. Une durée bien courte d'ailleurs, si l'objectif est de suivre dans le temps le ressenti du patient sous Kaletra.

3) Les enquêtes cas-témoins

Il s'agit d'enquêtes rétrospectives (basées sur des données recueillies dans le passé). Les enquêtes cas-témoins sont fréquentes, car elles ne coûtent pas cher. Un exemple-type : l'étude cas-témoins sur le lien entre vaccin anti-hépatite B et sclérose en plaques. On regarde la prévalence des événements démyélinisants chez les vaccinés versus chez les non vaccinés. Inconvénients : ces enquêtes sont très souvent biaisées et n'ont généralement pas une force de preuve très importante. En effet, il est difficile de trouver des témoins comparables en tout point aux cas, d'éviter les biais dû à un événement intercurrent, etc. Compte-tenu de leur faible valeur de preuve, ces enquêtes sont considérées comme une première étape indicative : si une enquête cas-témoin montre un lien entre deux événements, alors cela justifie des études complémentaires. Mais on ne peut généralement pas conclure sur la seule foi d'une enquête cas-témoins.

Principe de base : de manière générale, les études prospectives ont toujours une valeur de preuve plus forte que les études rétrospectives.

II - Description des données

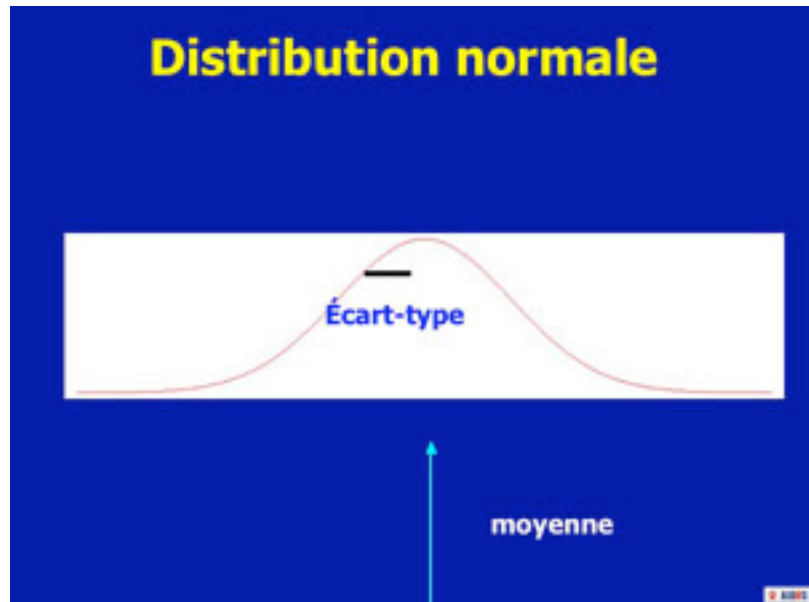
Les données recueillies dans le cadre des enquêtes peuvent être qualitatives ou quantitatives.

Données qualitatives

Ce sont des données en « classes ». Exemple : le sexe des répondants (homme/ femme/ trans).

- On peut présenter les données qualitatives en effectifs fréquences (n %).
- Pour le recueil, il faut éviter les questions ouvertes.
- Il est souvent commode de regrouper des données continues en classes (pour l'âge : de 15 à 25 ans, de 26 à 31 ans, etc.)
- Il faut faire le deuil de la représentation de toutes les situations dans le cadre de l'enquête : pour l'analyse, la représentation en classes est plus pratique,

même si elle ne permet pas de conclure à un résultat par âge donné.
De manière générale, on ne peut pas représenter avec exhaustivité l'ensemble des situations possibles. Pour l'analyse, il faut simplifier et regrouper les situations.



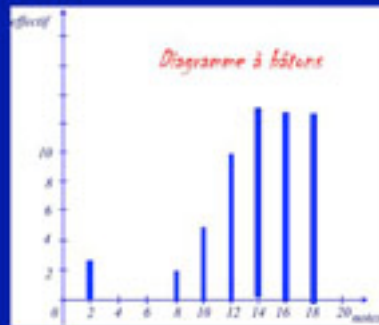
Données quantitatives

Exemple : le nombre de cigarettes fumées par jour, la charge virale, le nombre de CD4. Il s'agit de données continues. De même, on peut regrouper les données quantitatives par « classes » pour les représenter par exemple sous forme d'un camembert :

- ceux qui ne fument pas = une tranche ;
- ceux qui fument moins d'un paquet = une tranche ;
- ceux qui fument plus d'un paquet = une autre tranche.

Si les données quantitatives sont distribuées selon une loi statistique dite « normale » (i.e. suivant une courbe de Gauss = en forme de cloche, symétrique autour d'un pic) : on peut calculer la moyenne et l'écart-type. L'écart-type est significatif de la distribution des résultats autour de la moyenne i.e. plus l'écart-type est petit, plus les résultats sont groupés autour de la moyenne. Plus l'écart-type est grand, plus la distribution est hétérogène autour de la moyenne.

Distribution anormale



Médiane : partage de 50% de la distribution

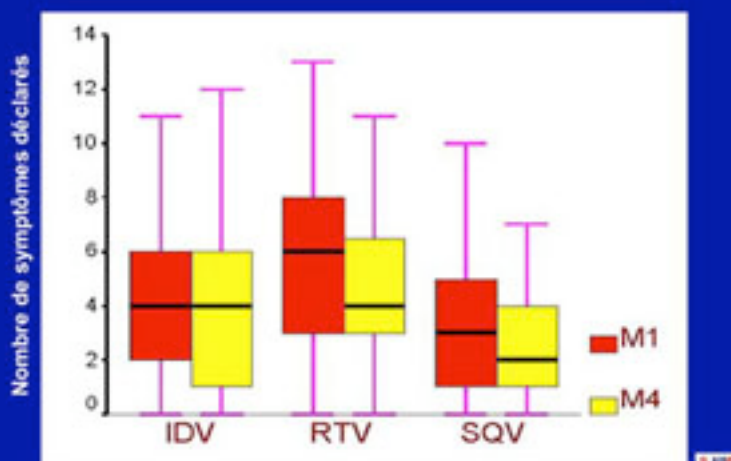
IQR: limites de 3/4 de la distribution

Rang: limites de la distribution entière

La plupart du temps, les données quantitatives recueillies ne suivent pas une loi statistique normale, mais dénotent une distribution anormale (i.e. qui n'est pas centrée). Les valeurs d'analyse utilisées dans le cas d'une distribution anormale sont :

- la médiane (elle partage les résultats en deux moitiés égales : il y a autant de données inférieures à la médiane que de données supérieures à la médiane ; la médiane n'est pas équivalente à la moyenne) ;
- l'écart interquartile (IQR) mesure la dispersion autour de la médiane. L'écart interquartile, c'est l'étendue de la distribution sur laquelle se trouve concentrée la moitié des éléments dont les valeurs sont les moins différentes de la médiane. On exclut de cette distribution 25% des valeurs les plus faibles et 25 % des valeurs les plus fortes ;
- le rang, qui correspond à l'écart observé entre les valeurs extrêmes recueillies.

Les boîtes à moustache



Exemple ci-contre : en ordonnée, le nombre d'effets indésirables rapportés. Avec l'indinavir (IDV) à M1, la médiane est de 4 : 50 % des répondants ont plus de 4 effets indésirables ; 50 % ont moins de 4 EI. Ecart interquartile (les limites de la boîte à moustaches) = 6 et 2. Les 50 % des répondants les plus proches de la médiane ont entre 6 et 2 EI. Le rang est égal à 0-11 : certains n'ont pas d'effet indésirable, d'autres en déclarent 11.

Avec le RTV à M1 :

- médiane : 6 EI.
- Ecart interquartile : entre 8 et 3 EI.
- Rang = 0-13.

III - Les analyses explicatives

L'objectif d'une analyse explicative est de rechercher les facteurs associés explicatifs d'une « variable réponse » de l'enquête (= équivalent du « critère principal de jugement » des essais cliniques). Les questions préalables à l'analyse :

- Quelle est la variable réponse ? (Exemple : il peut s'agir de l'observance au traitement)
- Comment se distribue-t-elle ? (Exemple : on peut avoir les catégories « très observants », « peu observants », « pas du tout observants »...)
- Quels sont les facteurs explicatifs possibles ? (Exemple : les conditions de logement ont-elles, oui ou non, une influence sur l'observance ?) Dans l'analyse explicative, on cherche les facteurs explicatifs (âge, sexe, etc.) et on évalue leur action sur la variable réponse.

Exemple : Si on veut évaluer l'action du sexe génétique sur l'observance, on va comparer des proportions (femme / homme) en fonction de la variable. On fait ensuite un calcul pour voir si les différences observées sont significatives ou non = la différence observée est-elle due au hasard ou non ?

Variable explicative	Observants	Non-Observants
% femmes	28	26
Age moyen	40	36

Dans l'exemple ci-dessus : la différence en fonction de l'âge est significative ; la différence en fonction du sexe ne l'est pas. Les malades observants sont significativement plus âgés que les malades non observants. Mais la proportion de femmes n'est pas significativement différente chez les observants et chez les non observants.

Les tests statistiques permettent de vérifier la significativité des résultats, i.e. de

dire si la différence observée est due au hasard ou non. Le test statistique est exprimé par « p » : « p » est le risque d'erreur associé au fait de dire qu'il existe une différence (on a décidé arbitrairement qu'on pouvait accepter une erreur inférieure à 5 %). Si $p < 0.05$, on décide que la différence observée est significative. $p < 0.05$ = le risque d'erreur est inférieur au seuil de 5 % choisi arbitrairement, on peut donc dire qu'il existe une différence. $p = 0.04$ signifie que l'on a 96 % de chances que la différence observée ne soit pas due au hasard, et 4 % de chances de se tromper. $p > 0.05$ = le risque d'erreur est supérieur à 5 %. On a plus de 5 % de chances de se tromper = la différence observée n'est pas significative. p est inférieur ou égal à 0.001 : on dit que la différence est « hautement significative » car le risque de se tromper est vraiment faible.

Expression par un odds ratio (OR) ou par le risque relatif (RR)

Exemple : risque de non-observance

Variable explicative	Observants	Non-Observants	OR [95%IC]
Femmes (%)	28	26	1.2 [0.90-1.56]
Sous IP (%)	58	66	1.5 [1.2-2.3]
Age moyen	40	36	0.96[0.94-0.99]

OR = 1,2 => Si vous êtes une femme, vous avez 1,2 fois plus de chances d'être moins observants que si vous êtes un homme. L'odds ratio (OR) exprime le risque d'être non observant si vous êtes une femme par rapport au risque d'être non observant si vous n'êtes pas une femme. La valeur de l'OR indique le sens de la relation entre le facteur (sexe) et la variable (observance), i.e. elle dit si le facteur est protecteur ou non, s'il favorise l'observance ou la non-observance par exemple. Pour savoir si l'association entre le facteur et la variable est significative, on doit regarder l'intervalle de confiance (IC). L'IC fait état de la marge d'erreur. Dans l'exemple, IC varie entre 0.90 et 1.56. Cela signifie que la valeur de l'OR a 95 % de chances d'être comprise entre 0.90 et 1.56. Attention une règle importante : si le 1 est dans l'intervalle de confiance, l'association entre le facteur et la variable n'est pas significative, puisque l'OR peut être associé à soit plus de risque, soit moins de risque. Dans l'exemple, l'intervalle de confiance varie entre 0.90 et 1.56. Le 1 est dans l'intervalle de confiance, l'OR n'est donc pas significatif, l'association entre le fait d'être une femme et le défaut d'observance n'est donc pas significative.

Dans l'exemple, pour le facteur explicatif « sous IP » : l'OR = 1.5 (IC compris entre 1.2 et 2.3) = c'est significatif. Le fait d'être sous IP est un facteur associé à une moins bonne observance.

La valeur absolue de l'OR a aussi une signification. Ex : OR = 4 (avec un intervalle de confiance significatif) = on peut en déduire que le lien entre le facteur et la

variable réponse est important.

Les limites des analyses univariées

Les petites analyses décrites ci-dessus sont des analyses univariées. Dans les analyses univariées, on se contente de regarder les odds ratios bruts, d'observer la force des associations, sans chercher à ajuster les résultats sur l'existence d'éventuelles variables de confusion. Ces analyses ont de nombreuses limites :

- la limite de « p » à 0.05 est arbitraire. Le sens de « p » s'interprète par rapport à la taille de l'échantillon : si « p » n'est pas très significatif mais que l'échantillon est petit, cela « augmente » la valeur du résultat. Si « p » n'est pas très significatif alors que l'échantillon est grand, c'est suspect ;
- dans ce type d'analyse, on constate une relation, mais pas une causalité ! Exemple : on peut voir une relation entre deux éléments (prendre une IP et être non observant), mais on ne peut pas affirmer la causalité car il peut y avoir des variables de confusion. Il faut ajuster les analyses par rapport aux possibles variables de confusion pour conclure à une causalité.

Exemple : on constate une relation entre le fait de manger une pizza et le fait de contracter une toxi-infection alimentaire. OR = 10 ; IC = 8 – 30. La relation est forte. Mais il y a une variable de confusion : le tiramisu. Pour sortir de la confusion, on fait une analyse en deux strates. On sépare deux groupes : ceux qui ont mangé du tiramisu et ceux qui n'en ont pas mangé. Puis on regarde les relations avec la toxi-infection dans les deux groupes. Plus aucune relation entre manger de la pizza et la toxi-infection n'est alors visible quand on stratifie par le fait d'avoir manger ou non du tiramisu, car c'est le tiramisu qui était associé à la toxi-infection et qui confondait le rôle de la pizza.

Voilà le principe de l'analyse multivariée : on intègre les variables de confusion à l'analyse univariée en faisant des groupes et les sous-analyses adéquates ; ainsi, on interprète mieux la relation observée.

L'analyse multivariée

L'objectif de l'analyse multivariée est de trouver les facteurs indépendamment associés à la variable réponse. Pour cela, on introduit tous les facteurs associés en analyse univariée à l'analyse multivariée, et on recherche les variables de confusion (qui peuvent être des facteurs associés). Les facteurs associés sont toujours le résultat d'hypothèses faites au niveau de l'analyse univariée. Exemple : le tiramisu est un facteur ajouté à l'analyse parce qu'une hypothèse a été formulée. Si on ne fait pas l'hypothèse du tiramisu, alors on peut avoir des résultats trompeurs. Dans de nombreuses enquêtes, l'âge peut être une variable de confusion toute bête. Il faut toujours raisonner à la manière des poupées russes pour ce type d'analyse : une variable peut en cacher une autre. Aujourd'hui, par de savants calculs automatiques, les logiciels de statistiques parviennent à faire disparaître les

facteurs les moins explicatifs et les plus confusionnants sans problème.

IV - Les modèles longitudinaux : quelques notions

1) Le modèle de Cox ou modèle de survie

Dans ce modèle, on va étudier un « premier événement » au cours du temps et modéliser les facteurs associés à ce premier événement. Exemple 1 : le premier événement peut être la progression clinique. On va compter les personnes qui passent en dessous de la barre des 200 CD4. Cf. diapo : on y compare le risque de progression clinique en fonction du fait d'être dépressif/ non dépressif. On observe que les dépressifs progressent plus vite dans la maladie que les non dépressifs. Dans l'exemple, avec le modèle de Cox, on recherche pourquoi les gens progressent :

- la dépression ;
- le % de non adhérence ;
- les CD4 à l'initiation.

Ces 3 facteurs sont indépendamment associés au risque de progresser, c'est-à-dire que même si les dépressifs prennent leur traitement, le simple fait qu'ils vivent une dépression augmente leur risque de progresser cliniquement.

2) Le Modèle GEE (generalised estimated equation)

Dans ce modèle, on fait l'hypothèse que le temps n'influence pas ce qu'on veut expliquer. On effectue des observations répétées au cours du temps.

3) Le modèle mixte

On effectue des observations répétées au cours du temps ; le temps influence les facteurs qui expliquent la variable réponse i.e. faire un événement à 2 ans n'a pas le même poids que faire un même événement à 3 ans...

4) Le modèle joint : il s'agit d'un mélange de plusieurs modèles.

V – Explication de la présentation réalisée pour le séminaire clinique de l'ANRS

Au séminaire de l'ANRS, Bruno a réalisé une présentation sur l'observance étudiée dans le cadre d'APROCO, dont nous avons compris le début, mais pas la fin. Nous lui demandons donc de nous expliquer son analyse.

Rappels sur APROCO :

- il s'agit d'une cohorte visant à suivre des patients VIH+ ayant initié une multithérapie comportant une IP ;
- c'est une cohorte nationale, qui implique de nombreux centres cliniques ;
- il est prévu un suivi à M0, M1, M4, M12 puis tous les 8 mois ;
- on y étudie des variables cliniques, biologiques, socio-comportementales (observance, dépression, effets indésirables perçus par les patients). Un auto-questionnaire d'observance est théoriquement remis à tous les patients à chaque visite.

On connaît désormais bien les facteurs qui influencent (en bien ou en mal) l'observance au traitement durant la phase d'induction. En revanche, les facteurs jouant sur l'observance au long cours sont moins documentés. Sur le long terme, les données manquantes dans les enquêtes deviennent un problème majeur : il y a l'attrition (les perdus de vue) + les personnes qui ne répondent pas aux auto-questionnaires pour diverses raisons (souvent, les médecins oublient eux-mêmes qu'ils doivent remettre un auto-questionnaire). Or, si on fait l'analyse avec les seules données des personnes qui ont rempli les auto-questionnaires, on a un sérieux biais, une population de personnes qui n'est pas représentative des personnes suivies.

L'objectif de l'étude présentée par Bruno au séminaire de l'ANRS était donc :

- d'observer les facteurs associés à la non observance pendant la phase de maintenance du traitement avec les données disponibles au long cours ;
- de vérifier la stabilité des résultats lorsqu'on tient compte du biais induit par les données manquantes.

La technique de calcul employée a fait appel aux compétences de nombreux spécialistes. C'est une adaptation de la méthode de Shelton, qui lui-même a adapté une méthode de Heckman...

La technique est basée sur une hypothèse. Cette hypothèse, c'est que la population des personnes qui n'a pas répondu aux autoquestionnaires sur le long cours est significativement différente de celle qui a répondu. On fait alors une analyse en deux stades :

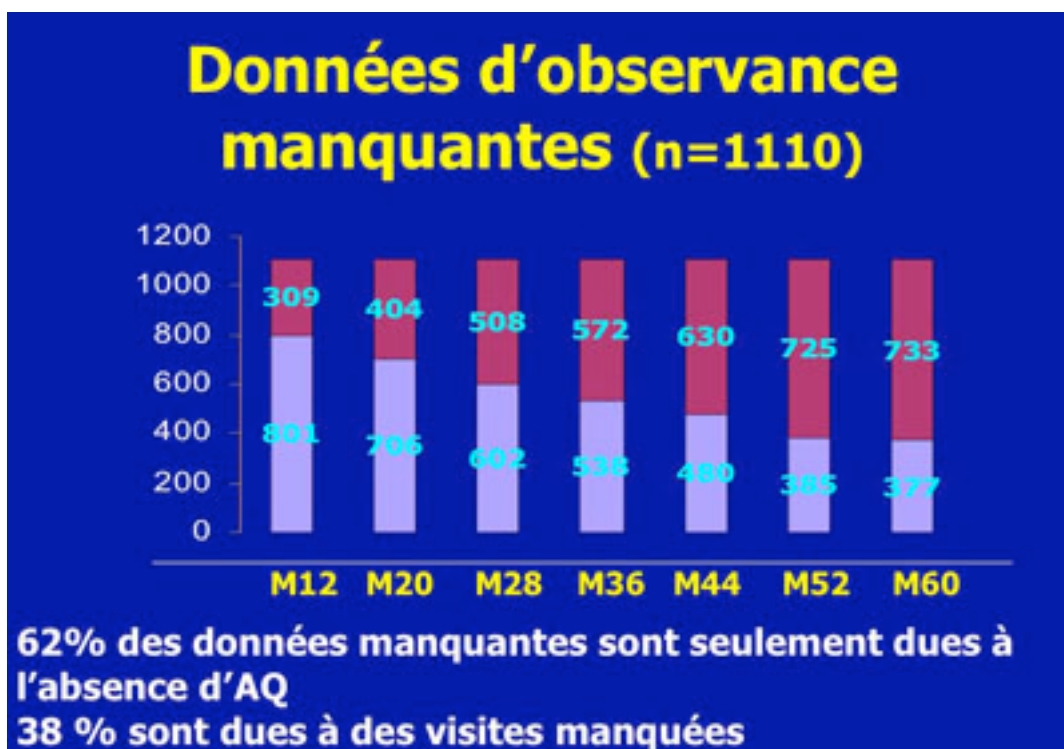
- au premier stade, on regarde pourquoi on a des données d'observance pour certaines personnes et pas pour d'autres ;
- dans le second stade, on regarde pourquoi les personnes sont observantes ou non. On utilise ensuite les résultats du premier stade pour pondérer le second stade.

Premier stade : on cherche les facteurs associés au remplissage et au non remplissage de l'autoquestionnaire. Par une équation, on modélise la probabilité d'avoir une donnée d'observance ou non en fonction des différents facteurs associés. Au bout de l'équation, il demeure toujours un « résidu », i.e. une part de « non réponse » qui n'est pas expliquée par les différents facteurs. On se sert de ce

résidu pour créer une nouvelle variable, appelée « la variable IMR ». Cette variable exprime la variabilité non expliquée du processus de non-réponse au questionnaire d'observance.

Selon notre hypothèse initiale, les non répondants sont significativement différents des répondants. Au deuxième stade de l'analyse, la variable IMR est utilisée comme variable d'ajustement du biais dû aux données manquantes. Si l'introduction dans l'analyse de la variable IMR ne modifie pas les résultats, alors cela signifie que notre hypothèse initiale était fausse (les non répondants ne sont pas significativement différents des répondants).

Mise en pratique : au premier stade de l'analyse, on regarde toutes les visites prévues de M12 à M60 et on note l'absence ou la présence du patient (total : 7 994 visites). Au deuxième stade, on regarde toutes les visites de M12 à M60 et on note si le questionnaire d'observance a été ou non rempli (seulement 3889 auto-questionnaires ont été remplis). On va ainsi faire la différence entre les personnes qui ne viennent pas du tout aux visites et celles qui viennent sans remplir l'auto-questionnaire.



Le tableau montre l'évolution du recueil des données au cours du temps, à M12, M20, M28, M36, M44, M52, M60. En bleu, les auto-questionnaires remplis ; en rouge, les questionnaires non remplis. D'après les premières analyses menées, on sait que, dans 62 % des cas, la non réponse au questionnaire d'observance n'est due qu'à l'absence de questionnaire : i.e. l'auto-questionnaire n'est pas parvenu jusqu'au patient parce que le médecin ne le lui a pas remis... Le questionnaire non

renseigné est associé à une visite manquée à l'hôpital dans seulement 38 % des cas.

Dans la suite de l'analyse, on établit la liste des facteurs associés au non remplissage du questionnaire et on modélise la non réponse au questionnaire par une équation. Comme prévu, il demeure un faible résidu, i.e. une part de non réponse qui n'est pas expliquée par l'ensemble des facteurs connus.

Exemple : le fait de ne pas répondre au questionnaire = X fois le fait d'avoir des effets indésirables + Y fois le fait de ne pas avoir de soutien dans son entourage + Z fois... etc. + résidu.

On va ensuite faire une analyse multivariée pour connaître les facteurs expliquant l'observance / la non observance au traitement.

Les résultats obtenus sont présentés sur la diapo n°13 (ci-dessous).

Modèle I sans IMR :

- cette fois, c'est le 0 qui ne doit pas être dans l'intervalle de confiance pour que le résultat soit significatif ;
- si l'odds ratio (coefficient dans le tableau) est positif, cela signifie qu'un facteur favorise un défaut d'observance ;
- si l'odds ratio est négatif, cela signifie que le facteur favorise l'observance.

Diapo 13 : facteurs associés à la non observance. Age élevé : plus on est âgé, plus on est observant. UDVI : moins bonne observance Prendre un traitement en une fois ou en 3 fois par jour est associé à une moins bonne observance.

Facteurs associés à la non-observance						
	Modèle I sans IMR			Modèle II avec IMR*		
	Coeff	95% IC		Coeff	95% IC	
Age	-0.03	-0.04	-0.02	-0.04	-0.04	-0.03
Bonnes conditions de logement	-0.11	-0.22	-0.00			
Né en France ou UE				0.16	-0.00	0.32
UDVI	0.18	0.00	0.36			
Naïf à M0				-0.19	-0.32	-0.06
IP dans la combinaison	0.17	0.05	0.28	0.16	0.05	0.28
Nombre de prises						
Qd vs. bid	0.45	0.17	0.74	0.47	0.18	0.75
Tid (ou plus) vs. bid	0.18	0.07	0.29	0.18	0.07	0.29
Nombre d'effets secondaires perçus	0.03	0.01	0.04	0.03	0.01	0.04
Dépression	0.19	0.08	0.30	0.18	0.07	0.29
Soutien élevé du partenaire	-0.15	-0.25	-0.05	-0.15	-0.25	-0.05
Diagnostic HIV < 6 mois	-0.15	-0.29	-0.01			
* l'IMR est hautement significatif						

A ce stade de l'analyse, on n'a pas pris en compte la variable IMR.

Modèle II avec IMR : Dans le modèle I, on a introduit tous les facteurs qui pouvaient expliquer la réponse/ non réponse au questionnaire d'observance et on les a testés sur l'observance au traitement ARV. Dans le modèle II, on va ajouter la variable IMR à l'analyse, qui correspond à un biais de présence ou non des données d'observance. L'ajout de l'IMR change les résultats de la première analyse i.e. certains facteurs qui étaient associés significativement à l'observance/ la non observance vont disparaître : le groupe de transmission (UDVI) et la situation sociale (conditions de logement). Cela signifie que les personnes contaminées par usage de drogues ont la même observance que les autres lorsqu'elles viennent en visite et qu'on leur donne l'occasion de remplir l'auto-questionnaire. L'introduction de la variable IMR modifie les résultats de l'analyse : cela signifie que l'IMR est hautement significatif et que l'hypothèse de départ était bonne. Les personnes qui n'ont pas rempli l'auto-questionnaire sont significativement différentes des personnes qui l'ont rempli.

On constate dans ce modèle II que les facteurs psychologiques ou thérapeutiques et leur impact sur l'observance ne sont pas modifiés par l'IMR. En revanche, les facteurs sociaux changent. Les grands messages restent inchangés :

- les patients ayant un niveau de vie modeste à faible ou les anciens UDVI ne sont pas moins observants que les autres dans la mesure où ils ont accès aux soins ;
- les migrants sont observants, plus observants même que les personnes nées en France, lorsqu'ils ont un bon accès aux soins. Cette analyse nous montre que, dans les enquêtes d'observance au long cours, on a une sur-représentation des personnes ayant un haut niveau social, qui n'ont pas été contaminées par usage de drogues. En revanche, des catégories sous-représentées sont les patients non migrants et non observants. Une mauvaise tolérance du traitement perçue par le patient favorise un défaut d'observance. Avoir un traitement en 3 prises ou en une prise par jour favorise le défaut d'observance, par rapport à un traitement en deux prises par jour.

Compte rendu rédigé par CT le 15 juillet 2005.