

Remerciements

Je voudrais remercier ici tous ceux qui m'ont aidé et soutenu durant l'ensemble de cette thèse, mes deux rapporteurs : Messieurs Bruno Falissard et Francis Guillemin et mes deux examinateurs : Madame Natacha Heutte et Monsieur Alain Leplege.

Je remercie très chaleureusement Madame Véronique Sébille et Monsieur Jean-Benoit Hardouin qui m'ont tous deux encadré avec beaucoup d'humanité, de générosité et de rigueur. J'espère sincèrement que cette thèse n'est que la première des multiples aventures que nous allons pouvoir partager.

Je ne remercierai jamais assez Sophie, qui m'a supporté, soutenu et encouragé tout au long de ma thèse. Merci mille fois de m'avoir relu et corrigé comme tu l'as fait !

de ces paramètres estimés est regroupé dans le vecteur $\hat{\phi}$.

En se basant sur ces paramètres estimés, il est possible de calculer les probabilités attendues de chacun des patterns de réponses. Le vecteur $\pi(\hat{\phi})$ est alors défini comme le vecteur des probabilités de chacun des patterns de réponse possibles, calculés à partir des paramètres estimés $\hat{\phi}$. Soit un pattern de réponses donné. Ce pattern est constitué de trois réponses : l_1 pour l'item 1, l_2 pour l'item 2 et l_3 pour l'item 3. Dans un premier temps, les probabilités d'observer la modalité de réponse l_1 à l'item 1 : $P(X_1=l_1|\hat{\phi})$, la modalité de réponse l_2 à l'item 2 : $P(X_2=l_2|\hat{\phi})$ et la modalité de réponse l_3 à l'item 3 : $P(X_3=l_3|\hat{\phi})$ sont calculées à partir d'un PCM en utilisant les valeurs estimées des paramètres : $\hat{\phi}$. Dans un second temps, grâce à la propriété d'indépendance locale du modèle utilisé, la probabilité d'observer le pattern de réponses $\{l_1, l_2, l_3\}$ est calculé comme le produit des probabilités d'observer chacune de ces réponses : $P(\mathbf{X}=\{l_1, l_2, l_3\}|\hat{\phi})=\prod_{j=1}^3 P(X_j=l_j|\hat{\phi})$.

Pour la réalisation des tests d'adéquation, on définit le vecteur \mathbf{p} comme le vecteur des probabilités observées de chacun des patterns de réponses possibles, calculées à partir de la répartition observée des réponses aux items.

Enfin, la matrice de contraste \mathbf{U} est calculée comme présenté ci-dessus. Ainsi, il est possible de calculer la fonction $\mathbf{d} : \mathbf{d}=\sqrt{n}\mathbf{U}'(\mathbf{p}-\pi(\hat{\phi}))$, la matrice $\hat{\mathbf{D}}_{\pi}$ (matrice diagonale des éléments de $\pi(\hat{\phi})$), la matrice de poids $\mathbf{W} : \mathbf{W}=\mathbf{U}'\hat{\mathbf{D}}_{\pi}\mathbf{U}$ et finalement la statistique de test $Q : Q=\mathbf{d}'\mathbf{W}^{-1}\mathbf{d}$, asymptotiquement distribué selon une loi du Chi-deux à $rangs(\mathbf{W})-5-1$ degrés de liberté.

5.1.3. Programme additionnel Stata® « *pcmodel* »

Deux programmes additionnels Stata® ont été réalisés : « *pcmodel* » [104] et « *pcmtest* » [105], permettant à la fois la réalisation de modèles de type PCM et RSM et de tests d'adéquation de ces modèles aux données effectivement observées. Le programme additionnel « *pcmodel* » permet la réalisation de modèles à effets aléatoires de type PCM ou RSM, en intégrant la possibilité d'introduire des covariables de groupe possiblement associées au trait latent évalué. Ce programme permet donc directement la réalisation de la méthode « Wald » proposée dans ce mémoire.

La syntaxe de ce programme est la suivante (les différentes options sont détaillées tableau 15) :

`. pcmodel liste_des_items [if], [categorical(liste_des_covariables)]`

continuous(*liste_des_covariables*) difficultés(*liste_des_noms_de_matrice*) iterate(#) adapt
robust from(*nom_de_matrice*) rsm estimateonly]

Ce programme permet à la fois l'estimation des paramètres d'un modèle de type PCM ou RSM, mais aussi le calcul d'indices facilitant son interprétation. Différents indices sont proposés.

- Un premier indice permet de fournir une approximation du coefficient de détermination du modèle par l'estimation du pseudo R^2 marginal de McFadden [109]. Ce pseudo R^2 est défini à partir des vraisemblances marginales du modèle réalisé : $V_m(modèle)$ et du modèle nul (défini comme un modèle logistique multinomial à risques adjacents avec pour seule covariable un effet aléatoire correspondant à l'effet individu, supposé suivre une loi normale de moyenne μ et de variance σ^2 – voir équation n°43) : $V_m(nul)$. Son calcul est détaillé à l'équation n°44.

$$\ln \left(\frac{P(X_{ij}=l|\mu, \sigma^2)}{P(X_{ij}=l-1|\mu, \sigma^2)} \right) = \theta \quad \theta \sim N(\mu, \sigma^2) \quad (43)$$

$$Pseudo R^2_{McFadden} = 1 - \frac{\ln(\hat{V}_m(modèle))}{\ln(\hat{V}_m(nul))} \quad (44)$$

- Lorsque des covariables de groupe sont intégrées dans le modèle, d'autres indices sont calculés, correspondant au pourcentage de variance du trait latent expliqué par chacune d'elles. Ces indices doivent être considérés comme une aide apportée à l'interprétation des coefficients associés à chacune des covariables introduites lorsque ces covariables sont significativement associées au trait latent considéré. Des covariables peuvent en effet être significativement associées à une variation du trait latent, mais expliquer une part plus ou moins importante de la variabilité de celui-ci.
- Ce calcul de pourcentage de variance expliquée peut être effectué par le biais de l'estimation des sommes des carrés de type III associées à chacune de ces covariables. Soit un modèle dans lequel 3 covariables suspectées d'influer sur la valeur du trait latent ont été introduites : les covariables A , B et C . La somme des carrés de type III associée à la covariable A peut être facilement estimée au moyen de modèles emboîtés : un premier contenant l'ensemble des covariables A , B et C (le modèle « complet ») et un second comprenant l'ensemble des covariables initiales sauf celle étudiée (ce modèle, appelé modèle « réduit », comportera dans ce cas uniquement les covariables B et C). La somme des carrés de type III associée à la covariable A est alors calculée comme la différence entre les sommes des carrés résiduels

du modèle « réduit » et du modèle « complet ». La proportion de variance du trait latent expliquée par l'introduction de la covariable A se calcule comme le rapport entre la somme des carrés de type III et la somme des carrés résiduels du modèle « réduit ».

5.1.4. Programme additionnel Stata® « *pcmtest* »

Le programme additionnel « *pcmtest* » [105] permet la réalisation de tests d'adéquation de modèles à effets aléatoires de type PCM ou RSM aux données analysées. Ce programme ne peut être utilisé qu'après avoir utilisé le programme « *pcmodel* », afin d'utiliser les estimations de paramètres fournies par ce dernier pour la réalisation des tests d'adéquation.

La syntaxe du programme « *pcmtest* » est la suivante (les différentes options sont détaillées tableau 16) :

. **pcmtest** [, group (partition_*des_scores*) nfit(#) power(#) alpha(#) graphics]

Ce programme permet la réalisation des tests R1m et Si, ainsi que l'estimation de la puissance *a posteriori* qui leur est associée.

L'estimation de cette puissance *a posteriori* se base sur une distribution non-centrée du Chi-deux : sous l'hypothèse nulle, les statistiques des tests R1m et Si sont asymptotiquement distribuées selon une loi du Chi-deux. Au contraire et sous l'hypothèse alternative, ces statistiques sont distribuées selon une loi du Chi-deux non centrée avec comme paramètre de décentralisation la statistique calculée du Chi-deux. Il est alors possible d'estimer la puissance *a posteriori* des tests R1m et Si au moyen de la fonction de répartition d'une telle distribution non-centrée du Chi-deux [50, 110].

Soit λ le paramètre de décentralisation, ξ le nombre de degrés de liberté et χ^2_{seuil} le seuil de rejet de l'hypothèse nulle au risque de première espèce égal à α . Soit $F\chi^2_{\lambda, \xi}$ la fonction de répartition d'une loi du Chi-deux de paramètre de décentralisation égal à λ et à ξ degrés de liberté. La puissance *a posteriori* d'un test d'adéquation effectué au risque de premier espèce fixé à α peut être calculée comme égale à $1 - F\chi^2_{(\lambda, \xi)}(\chi^2_{\text{seuil}})$.

L'intérêt d'une estimation *a posteriori* de la puissance d'un test d'adéquation est de pouvoir permettre d'en faciliter l'interprétation. Par exemple : si ce test d'adéquation est effectué sur un échantillon de très grande taille, le rejet de l'hypothèse nulle peut être lié soit à l'existence d'une

inadéquation réellement importante du modèle aux données, soit à l'existence d'une inadéquation d'importance minimale, mais rendue statistiquement significative de part la taille trop importante de l'échantillon utilisé (cette dernière situation pouvant être apparentée à un problème de sur-puissance).

Il peut alors apparaître intéressant d'envisager la réalisation du test d'adéquation non pas uniquement sur l'échantillon observé – dont l'effectif n_{Obs} conduit à une puissance du test non maîtrisée, mais possiblement aussi sur un échantillon virtuel présentant strictement la même distribution de patterns de réponses que l'échantillon effectivement observé mais dont l'effectif serait déterminé de façon à ce que la puissance *a posteriori* du test réalisé soit conforme à celle souhaitée par le chercheur. Sous l'hypothèse d'invariance de la distribution observée des réponses aux items, il est possible d'estimer la puissance *a posteriori* d'un test d'adéquation réalisé sur un échantillon virtuel d'effectif n_{Virt} possiblement différent de n_{Obs} , ou de façon symétrique de déterminer l'effectif nécessaire pour observer une puissance *a posteriori* donnée (ces deux possibilités étant proposées dans le programme additionnel « `pcmtest` ») : comme les statistiques des tests d'adéquation R1m et Si correspondent à des statistiques du Chi-deux, leur estimation pour un échantillon d'effectif n_{Virt} différent de celui de l'échantillon observé n_{Obs} mais présentant strictement la même distribution des patterns de réponses aux items correspond à l'estimation réalisée dans l'échantillon réellement observé, pondéré par le rapport entre l'effectif n_{Virt} et n_{Obs} .

5.1.5. Illustration de l'utilisation de « *pcmodel* » et « *pcmtest* »

Le jeu pathologique correspond à une addiction aux jeux de hasard et d'argent, pouvant s'accompagner de distorsions cognitives entraînant des représentations erronées de celui-ci. Celles-ci peuvent entraîner des modifications de la pratique de jeu des patients considérées, ceux-ci étant alors possiblement amenés à nier la part de hasard et de chance inhérente au jeu [111]. L'identification des distorsions cognitives pourrait permettre une meilleure prise en charge des joueurs pathologiques en proposant des modifications de croyances irrationnelles liées au jeu [112]. Par ailleurs, certaines études ont pu montrer que les distorsions cognitives pouvaient être influencées par le sexe des joueurs [113, 114], le type de jeu pratiqué (jeu de hasard – par exemple le loto, jeu de hasard et quasi-adresse – par exemple un jeu de paris sportifs, ou jeu de hasard et adresse – par exemple le poker [115]) ou le degré de pathologie du joueur [116].

Le questionnaire GABS (Gambling Attitudes and Beliefs Scale) [117] est un questionnaire composé de 35 items, développé pour évaluer les distorsions cognitives chez des personnes jouant de manière régulière. Dans sa version française, ce questionnaire est composé de 5 sous-dimensions traitant des distorsions cognitives concernant les stratégies de jeu (tactiques de jeu pour augmenter le gain, refus d'un arrêt du jeu, attitudes et comportement ayant valeur de fétiche), les superstitions (à propos de chance ou de malchance) et les ressentis émotionnels (sous forme d'excitation provoquée par le jeu) [118].

Dans cet exemple d'utilisation des modules « *pcmodel* » et « *pcmtest* », nous n'étudions que la sous-dimension mesurant l'excitation émotionnelle provoquée par le jeu. Cette sous-dimension est composée de 5 items (les items 1, 18, 26, 27 et 35, correspondant aux variables « *item1* », « *item18* », « *item26* », « *item27* » et « *item35* » de la base de données analysée). Chacun de ces items comporte 4 modalités de réponse : « complètement d'accord », « plutôt d'accord », « plutôt pas d'accord » et « pas du tout d'accord ». Les individus interrogés étaient caractérisés par leur sexe (variable « *sexe* » codée 1 pour les hommes et 2 pour les femmes), le type de jeu auquel ils jouaient le plus volontiers (variable « *jeu* » à trois modalités de réponse, codées 1 pour les « jeux de hasard », 2 pour les « jeux de hasard et quasi-adresse » et 3 pour les « jeux de hasard et adresse ») et la qualification de leur type de jeu (variable « *joueurs* » à trois modalités de réponse codées 1 pour les joueurs non pathologiques, 2 pour les joueurs pathologiques non traités et 3 pour les joueurs pathologiques traités pour leur pratique de jeu).

La population étudiée était composée de 628 individus, dont 210 (33 %) étaient des femmes. 41 % des joueurs interrogés n'étaient pas des joueurs pathologiques, 26 % étaient des joueurs pathologiques non traités et 33 % étaient des joueurs pathologiques sous traitement pour leur pratique de jeux. Les types de jeu pratiqués étaient pour 57 % des jeux de hasard, pour 30 % des jeux de hasard et quasi-adresse et pour 13 % des jeux de hasard et adresse. Seuls 28 individus n'avaient pas renseigné l'intégralité des items qui leur étaient soumis.

Tout d'abord, un modèle de type PCM comportant les trois covariables étudiées a été réalisé au moyen de la commande « *pcmodel* » :

```
. pcmodel item1 item18 item26 item27 item35, categorical(sexe jeu joueurs)
```

Ensuite, un test d'adéquation du modèle utilisé aux données à analyser a pu être réalisé au moyen de

la commande « *pcmtest* ». La partition des scores nécessaire à la réalisation de ce test a été réalisée de telle façon à regrouper les individus en groupes d'au moins 60 individus. Le regroupement des scores était le suivant : 0, 1-3, 4-5, 6-7, 8-9 et 10-15. La puissance souhaitée du test d'adéquation réalisé était de 95 % :

```
. pcmtest, power(0.95) group(0 3 5 7 9 15)
```

Nous présentons tout d'abord les résultats du test d'adéquation :

```
Global tests of the fit: test R1m
                       groups: 0 3 5 7 9 15
                       Number of individuals with missing data: 28 (4.46%)
```

	df	N = 600			N = 112		
		R1m	p-val	Power	R1m	p-val	Power
R1m	69	262.7	0.0000	1.0000	48.9	0.9682	0.9500

```
Items specific tests of the fit : tests Si
```

Item	df	N = 600			N = 112		
		Si	p-val	Power	Si	p-val	Power
gabs1 :	15	66.2	0.0000	1.0000	12.3	0.6553	0.5662
gabs18 :	15	128.7	0.0000	1.0000	23.9	0.0660	0.9055
gabs26 :	15	81.3	0.0000	1.0000	15.1	0.4421	0.6816
gabs27 :	15	120.9	0.0000	1.0000	22.5	0.0955	0.8819
gabs35 :	15	80.9	0.0000	1.0000	15.1	0.4471	0.6790

Chacun des tests réalisés (tant le test R1m que les tests Si) sur la population d'individus ayant répondu à l'intégralité des items (soit 600 individus) témoignait d'une mauvaise adéquation du modèle utilisé à la population de l'étude ($p < 10^{-5}$). Cependant, la puissance estimée *a posteriori* de ces tests était très importante, avoisinant les 100 %. On pouvait supposer que la taille importante de l'échantillon, entraînant une possible sur-puissance, permettait une mise en évidence de problèmes minimes d'adéquation entre le modèle et les données observées. L'effectif nécessaire pour réaliser un test R1m à la puissance de 95 % sur une population virtuelle dont la distribution des patterns de réponse aux items était strictement identique à celle de l'échantillon observé était de 112 patients. Lorsque les tests étaient réalisés sur cet échantillon virtuel, plus aucun d'eux ne mettait en évidence

un problème adéquat du modèle aux données analysées. Il était alors légitime d'interpréter les résultats de l'analyse des données à l'aide d'un modèle de type PCM.

Nous présentons maintenant les résultats du PCM obtenus par la commande « *pcmodel* » :

```
log likelihood: -3598.5302
Condition number: 1.1689
Number of individuals: 628
Number of items: 5
Number of covariates: 3
```

Parameters of the Latent trait distribution:

```
Identifiability constraint: latent trait mean in the subgroup
    sexe: 1, jeu: 1, joueurs: 1: fixed to 0
Variance of the Latent trait: Sigma2= 0.5791 (SE: 0.0702)
```

Latent trait group effect²:

	Coef.	S.E.	z	P> z	[95% C.I.]	SS.III	df	V.expl.
-----						25.563	1	0.9%
sexe:								
sexe: 1	0			
sexe: 2	0.1691	0.0941	1.80	0.072	-0.0153 0.3535			
-----						5.226	2	0.2%
jeu:								
jeu: 1	0			
jeu: 2	-0.0081	0.0974	-0.08	0.934	-0.1990 0.1828			
jeu: 3	0.2860	0.1261	2.27	0.023	0.0388 0.5332			
-----						956.799	2	35.5%
joueurs:								
joueurs : 1	0			
joueurs : 2	1.0399	0.1002	10.38	0.000	0.8436 1.2362			
joueurs : 3	1.2600	0.1012	12.45	0.000	1.0617 1.4583			
-----						SS.res	df	
						Model without covariates	2693.980	2983
						Full model	1724.453	2978

Items difficulty parameters³:

Item	Coef.	S.E.	[95% C.I.]

gabs1 :			

- 2 Pour rappel : la variable « sexe » est une variable dichotomique codée 1 pour les hommes et 2 pour les femmes. La variable « jeu » est une variable polytomique décrivant le type de jeu préféré des joueurs. Ses trois modalités de réponse sont codées 1 pour les « jeux de hasard », 2 pour les « jeux de hasard et quasi-adresse » et 3 pour les « jeux de hasard et adresse ». Enfin, la variable « joueurs » est une variable polytomique caractérisant les joueurs par rapport à leurs pratiques de jeu. Ses trois modalités de réponse sont codées 1 pour les joueurs non pathologiques, 2 pour les joueurs pathologiques non traités et 3 pour les joueurs pathologiques traités pour leur pratique de jeu
- 3 Pour rappel, chacun des items analysés comporte 4 modalités de réponse codés 1 pour « complètement d'accord », 2 pour « plutôt d'accord », 3 pour « plutôt pas d'accord » et 4 pour « pas du tout d'accord »

response :2	1.0775	0.1424	0.7984	1.3565
response :3	0.8342	0.1569	0.5266	1.1418
response :4	2.0803	0.1793	1.7288	2.4318

gabs18 :				
response :2	1.5351	0.1361	1.2683	1.8018
response :3	2.0057	0.1822	1.6486	2.3629
response :4	2.1530	0.2310	1.7002	2.6058

gabs26 :				
response :2	0.5308	0.1548	0.2275	0.8342
response :3	0.3056	0.1544	0.0030	0.6081
response :4	1.3092	0.1529	1.0095	1.6090

gabs27 :				
response :2	1.0611	0.1380	0.7907	1.3316
response :3	1.1761	0.1580	0.8665	1.4858
response :4	2.2241	0.1918	1.8481	2.6001

gabs35 :				
response :2	0.6122	0.1477	0.3226	0.9018
response :3	0.5988	0.1531	0.2986	0.8989
response :4	1.4830	0.1599	1.1696	1.7964

Ajusté sur le type de jeu et de joueurs, le sexe des patients n'était pas significativement associé à une variation d'excitation émotionnelle provoquée par le jeu. Par contre, une variation du type de jeu, tout comme une différence de type de joueurs, étaient significativement associées à une différence de niveau d'excitation émotionnelle : les personnes pratiquant des « jeux de hasard et adresse » présentaient un niveau d'excitation plus élevé que les autres joueurs. Une association statistique était bien mise en évidence, mais son importance pratique pouvait être relativisée, puisque la prise en compte du type de jeu ne permettait d'expliquer qu'à peine 0.2 % de la variance individuelle du score d'excitation émotionnelle.

Les caractéristiques des joueurs étaient elles aussi statistiquement associées au niveau d'excitation émotionnelle : les joueurs pathologiques (qu'ils soient traités ou non traités) présentaient un niveau d'excitation significativement plus élevé que les joueurs non pathologiques. Cette association semblait par contre présenter une réelle pertinence en pratique clinique, puisqu'elle permettait d'expliquer 35 % de la variabilité individuelle de l'excitation émotionnelle.

5.1.6. Intérêt, limites et perspectives

Les programmes additionnels « *pcmodel* » et « *pcmtest* » offrent certaines fonctionnalités

		Méthode « Wald »				Méthode « Wald-DIF »			
		Effet sexe	IC95%	p	Effet sexe	IC95%	p		
Dimensions physiques	Fonction physique (PF)	-0.21	-0.31	-0.11	0.000	0.01	-0.13	0.15	0.887
	Limitations physiques (RP)	-0.07	-0.21	0.07	0.314	-0.03	-0.17	0.11	0.687
	Douleurs physiques (BP)	-0.07	-0.14	-0.01	0.026	-0.05	-0.12	0.03	0.202
Dimensions physique / psychique	État de santé général (GH)	-0.16	-0.29	-0.04	0.008	-0.23	-0.36	-0.10	0.001
	Vitalité (VT)	-0.81	-1.09	-0.54	0.000	-0.33	-0.46	-0.19	0.000
Dimensions psychique	Vie de relation (SF)	-0.18	-0.24	-0.13	0.000	-0.18*	-0.24*	-0.13*	0.000*
	Santé psychique (MH)	-0.38	-0.49	-0.27	0.000	-0.27	-0.40	-0.14	0.000
	Limitations affectives (RE)	-0.17	-0.29	-0.04	0.011	-0.17	-0.30	-0.04	0.010

Tableau 24 : Modélisation de chacune des sous-dimensions de la qualité de vie évaluées à l'aide du questionnaire SF-36 en fonction du sexe, selon la méthode d'analyse choisie (méthode « Wald » ne prenant pas en compte le DIF, et méthode « Wald-DIF » prenant en compte un éventuel DIF). Seul l'effet du sexe sur la qualité de vie est reportée. Les paramètres de difficulté estimés à l'aide de chacun de ces modèles ne sont pas reportés.

IC95 % : intervalle de confiance à 95 % de l'estimation du paramètre associé à la covariable « sexe » dans le modèle considéré

p : valeur de significativité du paramètre associé à la covariable « sexe » dans le modèle considéré, estimée à l'aide d'un test de Wald.

Dimensions physiques : sous-dimensions du SF-36 fortement corrélées à la qualité de vie physique. Dimensions psychiques : sous-dimensions du SF-36 fortement corrélées à la qualité de vie psychique. Dimensions physiques / psychiques : sous-dimensions du SF-36 corrélées à la fois à la qualité de vie physique et psychique.

** : pas de DIF mis en évidence par la méthode « Wald-DIF ». Les résultats sont donc équivalents à ceux obtenus par la méthode « Wald ».*

		Méthode « Wald-DIF »	
		Items	Variation de difficulté
Dimensions physiques	Fonction physique (PF)	1, 2, 3, 4, 5, 6, 7	+ difficiles
	Limitations physiques (RP)	4	+ difficile
	Douleurs physiques (BP)	1	+ difficile
Dimensions physique / psychique	État de santé général (GH)	4	+ facile
	Vitalité (VT)	3, 4	+ difficiles
Dimensions psychique	Vie de relation (SF)	∅ DIF diagnostiqué	
	Santé psychique (MH)	2	+ difficile
	Limitations affectives (RE)	1	+ facile

Tableau 25 : Différence entre les valeurs des paramètres de difficultés associées aux différentes modalités de réponse des items de chaque sous-dimension observées chez les femmes et chez les hommes lorsqu'un éventuel phénomène de DIF lié au sexe est pris en compte lors de l'analyse du questionnaire SF-36 par la méthode « Wald-DIF ».

jusqu'alors non disponibles dans des logiciels statistiques, comme l'inclusion de covariables de groupes dans des modèles de type PCM ou RSM à effets aléatoires, le calcul d'indices facilitant l'interprétation des paramètres associés à ces covariables – comme le pourcentage de la variabilité du trait latent expliqué par l'introduction de chacune de ces covariables et la réalisation de tests d'adéquation – jusque-là réservés aux seuls modèles de Rasch ou aux modèles de type PCM ou RSM à effets fixes.

Cependant, ces programmes ne permettent la réalisation que des seuls tests R1m et Si et non du test R2m. Il n'est donc actuellement pas possible de tester l'ensemble des hypothèses inhérentes à la réalisation d'un modèle à crédits partiels, notamment l'hypothèse d'unidimensionnalité. Bien que cette dernière hypothèse soit habituellement vérifiée lors de la validation d'un questionnaire, il reste également nécessaire et souhaitable de la vérifier lors de l'étape d'analyse de données issues de ce questionnaire, notamment au moyen de méthodologies cohérentes avec l'ensemble de celles utilisées dans cette démarche. Une évolution prochaine du module « *pcmtest* » devrait combler cette lacune et permettre la réalisation d'un ensemble complet de tests d'adéquation dans le cadre de modèles de la famille de Rasch à effets aléatoires.

Enfin, il faut noter le fait que les tests d'adéquation ne peuvent être réalisés que sur l'ensemble des individus ayant répondu à l'ensemble des items. En effet, ces tests reposent sur le calcul des probabilités observées et attendues pour chaque pattern complet de réponses aux items. L'information disponible pour les individus n'ayant pas répondu à l'intégralité des items qui leur étaient soumis n'est donc utilisée que pour l'estimation des paramètres des modèles réalisés mais pas pour tester l'adéquation de ce modèle aux données observées.

5.2. Méthode « Two-Way » – Imputations multiples

5.2.1. Problématique

La méthode « Two-Way » telle que décrite dans ce mémoire repose sur l'estimation des paramètres d'un modèle d'ANOVA à 2 facteurs au moyen d'une procédure d'augmentation de données. Ce modèle d'ANOVA est utilisé dans un second temps pour imputer les données manquantes par un processus d'imputations multiples, permettant enfin l'analyse des données tant observées qu'imputées. Une telle procédure reste complexe à mettre en œuvre et de ce fait n'est pas couramment utilisée dans le domaine de l'analyse de questionnaires. À notre connaissance, un seul logiciel permet la réalisation d'imputations multiples en se basant sur un modèle d'ANOVA à effets mixtes. Il s'agit du programme « *tw.imputation* » issu de la librairie R « *miceadds* » [119]. Cette fonction n'est cependant pas intégrée dans un processus global et unifié d'analyse de questionnaires de type PROs : les imputations multiples des réponses manquantes aux items ne sont suivies ni d'un calcul de scores individuels issus de ces imputations ni d'une analyse aisée de ces derniers.

Le logiciel Stata® propose quant à lui un ensemble cohérent de fonctions regroupées sous l'environnement « *mi* » [120], permettant à la fois la réalisation d'imputations multiples de données manquantes et l'analyse de ces données imputées au moyen de méthodes appropriées. Cet environnement ne permet cependant pas la réalisation d'imputations multiples par un processus tel que celui proposé dans la méthode « Two-Way ».

Nous avons donc développé un programme additionnel Stata® : « *mi_twoway* » [121], permettant la réalisation d'imputations multiples basées sur un modèle d'ANOVA mixte dont les paramètres sont estimés par un processus d'augmentation de données et s'intégrant parfaitement à l'environnement « *mi* », permettant par là même une analyse facilitée des données imputées.

5.2.2. Estimation des paramètres d'un modèle mixte d'ANOVA par augmentation de données

Le modèle d'ANOVA sur lequel repose la méthode « Two-Way » est présenté ci dessus (voir page 62). Il définit la réponse observée d'un individu i à l'item j : X_{ij} comme étant fonction d'un effet « individu » : α_i (considéré comme étant un effet aléatoire distribué selon une loi normale de moyenne μ_{TW} et de variance σ^2_{TW}), d'un effet « item » : β_j (considéré comme étant un effet fixe) et d'une erreur de mesure (distribuée selon une loi normale de moyenne nulle et de variance σ^2_{TWres}) :

$$X_{ij} = \alpha_i + \beta_j + \varepsilon_{ij}.$$

L'estimation des paramètres de ce modèle ne peut être réalisée sans biais à partir des seules réponses observées si la répartition des données manquantes n'est pas parfaitement équilibrée

vis-à-vis des individus et des items [95]. Un processus d'augmentation de données [96] doit alors être mis en œuvre pour permettre une estimation non biaisée des paramètres recherchés. Ce processus de type « Espérance - Maximisation » a été décrit dans le cadre de données de type PROs par Van Ginkel [93] et est détaillé ci-dessous.

Soit un questionnaire comportant k items soumis à n individus. L'ensemble \mathbf{X} des réponses des n individus aux k items peut être décomposé en une partie observée : \mathbf{X}_{obs} et une partie manquante : \mathbf{X}_{mis} . L'ensemble des réponses observées est noté obs et l'ensemble des réponses manquantes est noté mis . Le nombre total de réponses observées est noté N_{obs} , le nombre total de réponses manquantes N_{mis} , le nombre total de réponses observées pour l'individu i $N_{obs(i)}$, le nombre total de réponses manquantes pour l'individu i $N_{mis(i)}$, le nombre totale de réponses observées pour l'item j $N_{obs(j)}$ et le nombre total de réponses manquantes pour l'item j $N_{mis(j)}$.

Les valeurs initiales du processus « Espérance - Maximisation » d'imputation des réponses manquantes : \widetilde{X}_{ij} sont initialement imputées par un tirage au sort aléatoire d'une valeur issue d'une loi normale de moyenne \hat{X}_{ij} et de variance S^2_{ini} (équations 45).

$$\hat{X}_{ij} = \sum_{j \in obs(i)} \frac{X_{ij}}{N_{obs(i)}} + \sum_{i \in obs(j)} \frac{X_{ij}}{N_{obs(j)}} + \sum_{i, j \in obs} \frac{X_{ij}}{N_{obs}}$$

$$S^2_{ini} = \sum_{i, j \in obs} \frac{(X_{ij} - \hat{X}_{ij})^2}{N_{obs} - 1}$$
(45)

Les valeurs initiales de μ_{TW} , σ^2_{TW} , σ^2_{TWres} , α_i et β_j , notés $\mu_{TW}^{(0)}$, $\sigma^2_{TW}^{(0)}$, $\sigma^2_{TWres}^{(0)}$, $\alpha_i^{(0)}$ et $\beta_j^{(0)}$, sont alors estimées de la façon suivante :

$$\mu_{TW}^{(0)} = \frac{\sum \sum_{i, j \in obs} X_{ij} + \sum \sum_{i, j \in mis} \widetilde{X}_{ij}}{n.k}$$

$$\beta_j^{(0)} = \frac{\sum_{i \in obs(j)} X_{ij} + \sum_{i \in mis(j)} \widetilde{X}_{ij}}{n} - \mu_{TW}^{(0)}$$

$$\alpha_i^{(0)} = \frac{\sum_{j \in obs(i)} X_{ij} + \sum_{j \in mis(i)} \widetilde{X}_{ij}}{k}$$

$$\sigma^2_{TWres}^{(0)} = \frac{\sum \sum_{i, j \in obs} (X_{ij} - \alpha_i^{(0)} - \beta_j^{(0)})^2 + \sum \sum_{i, j \in mis} (\widetilde{X}_{ij} - \alpha_i^{(0)} - \beta_j^{(0)})^2}{(n-1).(k-1)}$$

$$\sigma^2_{TW}^{(0)} = \sum_{i=1}^n \frac{(\alpha_i^{(0)} - \mu_{TW}^{(0)})^2}{n-1}$$
(46)

À l'itération t , les valeurs de μ_{TW} , σ^2_{TW} , σ^2_{TWres} , α_i et β_j , notés $\mu_{TW}^{(t)}$, $\sigma^2_{TW}^{(t)}$, $\sigma^2_{TWres}^{(t)}$, $\alpha_i^{(t)}$ et $\beta_j^{(t)}$, sont

estimées de la façon suivante :

- $\alpha_i^{(t)}$ est issu d'un tirage au sort aléatoire au sein de sa distribution postérieure conditionnelle aux estimations actuelles des autres paramètres, distribution de moyenne $\mu_{\alpha_i}^{(t)}$ et de variance

$S_{\alpha_i}^{2(t)}$:

$$\mu_{\alpha_i}^{(t)} = \frac{\frac{\mu_{TW}^{(t-1)}}{\sigma_{TW}^2(t-1)} + \sum_{j \in \text{obs}(i)} \frac{X_{ij} - \beta_j^{(t-1)}}{\sigma_{TW_{res}}^2(t-1)}}{\frac{1}{\sigma_{TW}^2(t-1)} + \frac{N_{\text{obs}(i)}}{\sigma_{TW_{res}}^2(t-1)}} \quad (47)$$

$$S_{\alpha_i}^{2(t)} = \frac{1}{\frac{1}{\sigma_{TW}^2(t-1)} + \frac{N_{\text{obs}(i)}}{\sigma_{TW_{res}}^2(t-1)}}$$

- $\beta_j^{(t)}$ est issu d'un tirage au sort aléatoire au sein de sa distribution postérieure conditionnelle aux estimations actuelles des autres paramètres, distribution de moyenne $\mu_{\beta_j}^{(t)}$ et de variance

$S_{\beta_j}^{2(t)}$:

$$\mu_{\beta_j}^{(t)} = \sum_{i \in \text{obs}(j)} \frac{X_{ij} - \alpha_i^{(t-1)}}{N_{\text{obs}(j)}} \quad (48)$$

$$S_{\beta_j}^{2(t)} = \frac{\sigma_{TW_{res}}^2(t-1)}{N_{\text{obs}(j)}}$$

- $\sigma_{TW_{res}}^2(t)$ est issu d'un tirage au sort aléatoire au sein de sa distribution postérieure conditionnelle aux estimations actuelles des autres paramètres, distribution inverse du Chi-deux par changement d'échelle à N_{obs} degrés de liberté et à paramètre d'échelle égal à $\Xi^{2(t)} = \sum \sum_{i,j \in \text{obs}} (X_{ij} - \alpha_i^{(t)} - \beta_j^{(t)})^2 / N_{\text{obs}}$. Ceci revient plus simplement à réaliser un tirage au sort au sein d'une distribution du Chi-deux à N_{obs} degrés de liberté et à attribuer à $\sigma_{TW_{res}}^2(t)$ la valeur transformée $\sigma_{TW_{res}}^2(t) = N_{\text{obs}} \cdot \Xi^{2(t)} / \chi_{N_{\text{obs}}}^2$.
- $\mu_{TW}^{(t)}$ est issu d'un tirage au sort aléatoire au sein de sa distribution postérieure conditionnelle aux estimations actuelles des autres paramètres, distribution de moyenne $\mu_{\mu_{TW}}^{(t)}$ et de variance

$S_{\mu_{TW}}^{2(t)}$:

$$\mu_{TW}^{(t)} = \sum_{i=1}^n \frac{\alpha_i^{(t)}}{n} \quad (49)$$

$$S_{TW}^{2(t)} = \frac{\sigma_{TW}^{2(t-1)}}{n}$$

- $\sigma_{TW}^{2(t)}$ est issu d'un tirage au sort aléatoire au sein de sa distribution postérieure conditionnelle aux estimations actuelles des autres paramètres, distribution inverse du Chi-deux par changement d'échelle à n degrés de liberté et à paramètre d'échelle égal à $\sum_{i=1}^n (\alpha_i^{(t)} - \mu_{TW}^{(t)})^2 / n$.

Le nombre d'itérations T est fixé *a priori*, permettant d'obtenir les estimations des paramètres du modèle pour un jeu de données imputées. Soit s le nombre d'imputations multiples souhaité. L'ensemble de ce processus (estimation de valeurs initiales, puis processus itératif de type « Espérance - Maximisation ») est répété s fois, créant s jeux de paramètres du modèle. Il est alors possible de vérifier la convergence de ces estimations au moyen de mesures de chaînes multiples [122], basées sur les comparaisons des variances des paramètres estimés sur l'ensemble des T . s itérations effectuées, et sur les T itérations effectuées pour chacune des s séries d'imputation.

5.2.3. Programme additionnel Stata® « *mi_twoway* »

Le programme « *mi_twoway* » [121] permet la réalisation d'imputations multiples basées sur une modélisation des réponses aux items soit par un modèle d'ANOVA à deux facteurs fixes, dont les paramètres sont estimés sur l'ensemble des cas observés sans procédure d'augmentation de données [94, 123], soit par un modèle mixte d'ANOVA à 2 facteurs dont les paramètres sont estimés via la procédure d'augmentation de données telle que présentée ci-dessus. Ce programme s'intègre parfaitement dans l'environnement « *mi* » proposé par Stata®, ainsi l'analyse des données imputées est directement accessible au moyen des différentes procédures de cet environnement, permettant la réalisation aisée de la méthode « Two-Way » proposée dans ce mémoire.

La syntaxe de ce programme est la suivante (les différentes options sont détaillées tableau 26) :

```
. mi_twoway liste_des_items [, scorename(nom_de_nouvelle_variable) replace add(##) style(dénomination_du_style) clear da iterate(##)]
```

Options	Description
scorename	Détermine le nom de la variable créée contenant les valeurs des scores individuels calculés comme la somme des réponses aux items. « scorename » est donc manquant pour chaque individu présentant au moins une réponse manquante à l'un des items qui lui étaient soumis. Par défaut, ce nom de variable est « score ».
replace	Permet de remplacer le contenu d'une variable existante dénommée « scorename ».
add	Détermine le nombre d'imputations multiples à effectuer. Par défaut, ce nombre est fixé à 10.
style	Détermine le style dans lequel doivent être présentées les données imputées. Différents styles sont proposés dans l'environnement « mi » : « wide », « mlong », « flong » et « flongsep ». L'ensemble de ces styles est utilisable avec « mi_twoway ».
clear	Permet de réaliser de nouvelles imputations multiples en supprimant celles précédemment réalisées.
da	Permet la réalisation d'imputations multiples basées sur une estimation par augmentation de données des paramètres d'un modèle d'ANOVA mixte à deux facteurs. Par défaut, les imputations sont réalisées par l'utilisation d'un modèle d'ANOVA à deux facteurs fixes.
iterate	Détermine le nombre d'itérations nécessaires à la réalisation de la procédure d'augmentation de données. Par défaut, ce nombre est fixé à 10.

Tableau 26: Liste des options du programme additionnel Stata « mi_twoway »

6. LIMITES POTENTIELLES – LE CAS DU FONCTIONNEMENT DIFFÉRENTIEL DES ITEMS

6.1.1. Présentation du fonctionnement différentiel des items (DIF)

Deux stratégies de comparaison de groupes d'individus basées sur l'analyse de PROs ont pu être identifiées comme étant particulièrement intéressantes en cas de données manquantes : les méthodes « Two-Way » et « Wald ». Ces stratégies permettent en effet d'obtenir des résultats non biaisés même lorsque les données manquantes sont informatives, c'est-à-dire liées à la valeur du trait latent des individus qui n'ont pas répondu à l'ensemble des items.

Cependant et pour aussi prometteuses qu'elles puissent paraître, ces méthodes présentent une limite commune. Celles-ci reposent sur l'hypothèse particulièrement forte selon laquelle les caractéristiques des items doivent être considérées comme étant identiques pour tous les individus, quelles que soient leurs caractéristiques socio-culturelles ou la période à laquelle ils sont interrogés.

Cette hypothèse peut à certains égards sembler peu crédible. La langue et ses structures linguistiques ne sont en effet pas les seuls à conditionner le langage. Celui-ci peut aussi être déterminé par des facteurs externes, comme des facteurs démographiques, sociaux ou économiques

[124, 125]. Ainsi, une simple différence de sexe peut se retrouver associée à des différences de langage [126, 127]. Une même idée pourra par exemple être exprimée au travers de mots différents selon qu'elle l'est par un homme ou une femme, réciproquement une même succession de mots être associée à des concepts différents. Ainsi, un même item pourra référer à deux questions distinctes selon qu'il est soumis à un homme ou une femme, ses propriétés psychométriques ne pouvant dès lors plus être considérés identiques pour l'ensemble des personnes interrogées.

On peut définir le fonctionnement différentiel d'un item (DIF) comme « une différence de performances à un item entre deux sous-groupes d'individus comparables entre eux, c'est-à-dire lorsque les sous-groupes sont appariés par rapport au concept hypothétique mesuré par le test. » [100]. Des différences langagières liées à des variations socio-culturelles, c'est-à-dire des sociolectes différents pourront être responsables de phénomènes de DIF : les différences de réponses aux items ne seront alors plus nécessairement le reflet d'une différence de valeur de trait latent, mais possiblement le reflet de variations d'interprétation de l'item considéré. Dans le cas d'un questionnaire dont l'objet est la mesure de la qualité de vie, une différence de scores pourrait alors tout aussi bien refléter une différence réelle de qualité de vie, qu'un phénomène de fonctionnement différentiel des items.

6.1.2. Méthode de prise en compte du DIF lors de l'analyse de PROs

La prise en compte d'un tel phénomène lors de l'analyse de PROs est délicate lorsque l'on choisit d'analyser ces réponses par le calcul de scores. En effet, l'algorithme permettant le calcul du score – le plus souvent distribué concomitamment au questionnaire – est identique pour chaque individu. La seule façon de prendre en compte un phénomène de DIF lors de l'analyse d'un questionnaire par l'intermédiaire du score serait de proposer un algorithme de calcul de ce dernier différent en fonction des groupes socio-culturels caractérisant les individus interrogés – par exemple un algorithme de calcul du score différent pour les hommes et les femmes, si on considère les différents sociolectes liés au genre.

Le diagnostic et la prise en compte de phénomènes de DIF peuvent s'avérer plus facilement envisageables à l'aide de modèles issus de la théorie de la réponse à l'item. Ainsi et dans des modèles de la famille de Rasch, il est possible d'intégrer des interactions entre les paramètres d'items et les groupes socio-culturels des personnes interrogées, permettant de considérer des

paramètres d'items dépendants de ces groupes d'appartenances.

Reprenons le cas d'un PCM incluant des covariables de groupe (modèle n°24)

$$P(X_{ij}=l|\boldsymbol{\beta}, \mathbf{c}_i, \theta_{Res}, \boldsymbol{\delta}_j) = \frac{\exp\left[l(\boldsymbol{\beta}'\mathbf{c}_i + \theta_{Res}) - \sum_{k=1}^l \delta_{jk}\right]}{\sum_{r=0}^{m_j} \exp\left[r(\boldsymbol{\beta}'\mathbf{c}_i + \theta_{Res}) - \sum_{k=1}^r \delta_{jk}\right]} \quad (50)$$

Avec un tel modèle, le trait latent étudié pouvait être influencé par l'ensemble des m covariables considérées. \mathbf{c}_i était le vecteur indicateur de dimension m spécifiant les valeurs observées des m covariables étudiées pour l'individu i et $\boldsymbol{\beta}$ le vecteur de dimension m des paramètres de régression associés aux m covariables, décrivant l'influence de ceux-ci sur le trait latent. On peut maintenant considérer que d'autres covariables au nombre de n (ces n covariables pouvant être identiques ou différentes des m covariables précédemment considérées) puissent être responsables de phénomènes de DIF, c'est-à-dire d'influencer les valeurs des paramètres de difficultés associés à chacun des items. Si \mathbf{d}_i est le vecteur indicateur de dimension n spécifiant pour chaque individu i les valeurs observées à ces n covariables, les valeurs des difficultés de chaque item j pourront être dépendantes des valeurs observées \mathbf{d}_i . Si on note \mathbf{v}_{jk} le vecteur de dimension n correspondant aux variations de difficulté de la $k^{\text{ième}}$ modalité de réponse de l'item j liées aux différentes valeurs des n covariables considérées et $\delta_{0_{jk}}$ la difficulté associée à la $k^{\text{ième}}$ modalité de réponse de l'item j lorsque l'ensemble de ces n covariables est égal à 0, la difficulté associée à la $k^{\text{ième}}$ modalité de réponse de l'item j , dépendante des n covariables étudiées, peut être écrite telle que : $\delta_{jk} = \mathbf{v}_{jk}'\mathbf{d}_i + \delta_{0_{jk}}$.

Ainsi, un modèle de type PCM, incluant des covariables de groupe pouvant influencer à la fois la valeur du trait latent, mais aussi les paramètres de difficulté associées aux items, peut être écrit tel que :

$$P(X_{ij}=l|\boldsymbol{\beta}, \mathbf{c}_i, \theta_{Res}, \boldsymbol{\delta}_j) = \frac{\exp\left[l(\boldsymbol{\beta}'\mathbf{c}_i + \theta_{Res}) - \sum_{k=1}^l (\mathbf{v}_{jk}'\mathbf{d}_i + \delta_{0_{jk}})\right]}{\sum_{r=0}^{m_j} \exp\left[r(\boldsymbol{\beta}'\mathbf{c}_i + \theta_{Res}) - \sum_{k=1}^r (\mathbf{v}_{jk}'\mathbf{d}_i + \delta_{0_{jk}})\right]} \quad (51)$$

La méthode « Wald » telle que présentée dans ce mémoire ne prenait pas en compte la possibilité de phénomènes de DIF. Cette méthode pourrait être adaptée au cas du fonctionnement différentiel des items en utilisant un modèle tel que celui présenté ci-dessus. Différents problèmes persistent cependant, pour connaître la meilleure façon de procéder à cette adaptation.

Le modèle présenté ci-dessus permet d'envisager un phénomène de DIF lié à un ensemble de covariables, possiblement présent sur l'ensemble des items du questionnaire et d'intensité variable selon les items et les modalités de réponse considérés. Cependant, rien n'implique que l'ensemble des covariables considérées soit effectivement responsable de phénomènes de DIF ni que celui-ci soit variable pour chacune des modalités de réponse de chaque item. Par exemple, si la variation de compréhension touche l'énoncé d'un item et non pas les modalités de réponse à celui-ci, le phénomène de DIF pourra être homogène pour l'ensemble de ces modalités de réponse. La problématique majeure liée à l'utilisation de ce modèle consiste donc à diagnostiquer le DIF, c'est-à-dire à identifier quelles covariables sont responsables d'un phénomène de DIF, quels items sont touchés par celui-ci et de quel type de DIF il s'agit. Autrement dit, la problématique majeure liée à l'utilisation de ce modèle consiste à déterminer le modèle réduit le plus pertinent, avant de chercher à étudier les propriétés d'un tel modèle.

Une méthode possible consiste à créer un modèle par procédure ascendante basée sur le critère d'Akaike. Le modèle initial correspond alors en un modèle minimal – c'est-à-dire un modèle de type PCM sans aucune covariable ni interaction. À chaque étape, un paramètre est ajouté au modèle – ce paramètre pouvant être soit une interaction entre l'un des items et une covariable (correspondant à un phénomène de DIF), soit à une simple covariable pouvant influencer sur le trait latent, permettant de minimiser le critère d'Akaike. La procédure s'arrête lorsque l'ajout d'un nouveau paramètre dans le modèle conduirait à une augmentation du critère d'Akaike.

6.1.3. Illustration de la méthode : étude de l'association entre sexe et qualité de vie chez des patients souffrant de lombalgie chronique

Une telle procédure peut être illustrée au moyen de l'étude du lien entre le sexe et la qualité de vie de patients souffrant de lombalgie chronique [128].

La lombalgie chronique est un problème majeur de santé publique. Ce symptôme touche environ

une personne sur quatre et est responsable d'une baisse très importante de la qualité de vie, plus importante que celle observée pour d'autres pathologies chroniques telles que le diabète, l'asthme ou l'hypertension. Dans le cadre de la lombalgie chronique, différents déterminants d'une baisse de qualité de vie ont pu être mis en évidence. Parmi ceux-ci, le sexe semble jouer un rôle majeur, les femmes présentant des scores de qualité de vie plus faibles que les hommes.

Il ne peut cependant être exclu que de telles différences ne soient pas uniquement le reflet d'une différence réelle de qualité de vie, mais possiblement d'un phénomène de DIF lié au sexe, c'est-à-dire d'une différence de compréhension, ou d'interprétation des items selon le sexe [129-131].

L'objectif de cette étude est donc d'étudier les relations entre le sexe et la qualité de vie de personnes souffrant de lombalgie chronique en prenant en compte lors de l'analyse un possible fonctionnement différentiel des items lié au sexe.

6.1.3.1. Méthodes

L'étude a porté sur 1342 personnes souffrant de lombalgie chronique. Ces personnes : 768 hommes et 574 femmes, étaient issues du réseau de surveillance des troubles musculo-squelettiques chez les salariés de la région Pays de la Loire, mis en place en 2002 par l'InVS. Leur qualité de vie était évaluée au moyen du questionnaire SF-36, composé de 8 sous-dimensions : les sous-dimensions « Fonction physique » (PF), « limitations physiques » (RP), « douleurs physiques » (BP), « vie de relation » (SF), « santé psychique » (MH), « limitations affectives » (RE), « état de santé général » (GH) et « vitalité » (VT). L'ensemble de ces sous-dimensions reflète à la fois la qualité de vie physique et psychique, chacune de ces sous-dimension étant plus ou moins corrélée avec la qualité de vie physique ou psychique. Ainsi, les sous-dimensions PF, RP et BP sont très fortement corrélées avec la qualité de vie physique, les sous-dimensions SF, MH et RE sont très fortement corrélées avec la qualité de vie psychique et les sous dimensions GH et VT sont corrélées à la fois avec la qualité de vie physique et psychique [132]. Chacune de ces 8 sous-dimensions a été analysée séparément. Deux stratégies d'analyse ont été suivies : une première ne prenait pas en compte un possible phénomène de DIF : il s'agissait de la méthode « Wald » telle que décrite dans ce mémoire, la covariable de groupe incluse dans le modèle correspondant alors au sexe des personnes interrogées. La seconde était basée sur la procédure ascendante telle que décrite ci-dessus, permettant d'envisager un éventuel DIF lié au sexe. Cette méthode est appelée par la suite méthode « Wald-DIF »

6.1.3.2. Résultats

Les données observées comportaient très peu de données manquantes : le taux de réponses manquantes était inférieur à 3 % pour chacun des items du SF-36.

La méthode « Wald » – interdisant la prise en compte d'un possible DIF lors de l'analyse – conduisait à la mise en évidence de différences significatives de qualité de vie tant physique que psychique entre les hommes et les femmes. Seule l'analyse de la sous-dimension RP ne conduisait pas à observer une différence significative liée au sexe (tableau 24).

Lorsqu'on permettait la prise en compte d'un éventuel DIF lors de l'analyse par la méthode « Wald-DIF », on n'observait plus de différence significative de qualité de vie pour les sous-dimensions purement physiques (c'est-à-dire les sous-dimensions fortement corrélées uniquement avec la qualité de vie physique : les sous-dimensions PF, RP et BP). Les seules différences persistantes concernaient les sous-dimensions associées uniquement ou en partie à la qualité de vie psychique. Pour l'ensemble des sous-dimensions traitant spécifiquement de la qualité de vie physique, les difficultés des items étaient plus importantes chez les femmes que chez les hommes. À qualité de vie identique, les femmes avaient systématiquement plus de difficulté à choisir des modalités de réponse décrivant positivement leur qualité de vie physique : celles-ci verbalisaient leur qualité de vie physique par des mots systématiquement plus péjoratifs que les hommes, sans pour autant présenter de différence significative de niveau de qualité de vie physique (tableau 25).

6.1.3.3. Discussion

L'analyse de la qualité de vie de patients souffrant de lombalgie chronique a conduit à des résultats radicalement différents selon que d'éventuels phénomènes de DIF sont ou non pris en compte. Lorsque ce phénomène n'était pas pris en compte, des différences de qualité de vie entre hommes et femmes étaient systématiquement observées, tant en ce qui concernait la qualité de vie physique que psychique. Par contre et lorsque cet éventuel phénomène était pris en compte, les seules différences persistantes de qualité de vie entre hommes et femmes correspondaient à des différences de qualité de vie d'ordre psychique. Ces différences psychiques entre hommes et femmes peuvent être mises en relation avec la plus grande prévalence de la dépression chez les femmes que chez les hommes. La variation d'interprétation des items, notamment concernant les items des sous-dimensions physiques peut quant à elle être expliquée par des différences culturelles liées au sexe, responsables d'un sociolecte spécifique parfois appelé « genrelecte » [133-135].

Les résultats de l'analyse présentée ci-dessus sont cependant à interpréter avec prudence : les caractéristiques des différentes méthodes d'analyse envisagées en présence de DIF ne sont pour l'instant pas connues. Il est possible qu'une méthode ne prenant pas en compte un éventuel phénomène de DIF (comme la méthode Wald) soit robuste vis-à-vis de ce dernier – c'est-à-dire que les résultats observés avec la méthode « Wald » concernant des différences de qualité de vie entre hommes et femmes soient exacts. Il est également tout à fait possible qu'une présence de DIF puisse induire des biais lors de l'estimation de différences de traits latents entre différents groupes d'individus par cette même méthode « Wald ». On ne sait si la méthode de diagnostic de phénomènes de DIF utilisée par la méthode « Wald-DIF » (basée sur le critère d'Akaike lors d'un processus ascendant de création de modèles) est une méthode fiable ou non de DIF, ni si les différences estimées de traits latents entre groupes d'individus estimées avec cette méthode sont ou non biaisées. Enfin, on ne sait pas si la complexification du modèle par l'utilisation de la méthode « Wald-DIF » induit une variation sensible de puissance par rapport à la méthode « Wald », empêchant une mise en évidence de phénomènes pourtant réels. Ainsi, cette étude ne peut actuellement avoir pour seule valeur que d'illustrer des différences possibles de résultats liées à des choix de stratégies d'analyse différents. Seules des études de simulation évaluant les propriétés de ces méthodes en présence de DIF permettront de déterminer quelle méthode devrait être utilisée dans quelle situation.

7. DISCUSSION

De plus en plus d'études de recherche clinique ont pour objet l'étude et l'analyse de mesures subjectives réalisées au moyen d'auto-questionnaires standardisés. L'analyse de ces questionnaires passe alors habituellement par le calcul de scores, alors que d'autres stratégies existent et peuvent être proposées, notamment basées sur l'utilisation de modèles de la famille de Rasch.

Au cours de ce mémoire, deux stratégies ont été sélectionnées parmi les 12 méthodes initialement

envisagées. La première, basée sur l'IRT, consistait en l'analyse d'une covariable de groupe incluse dans un modèle de la famille de Rasch à effets aléatoires (il s'agissait de la méthode « Wald »). La seconde, basée sur la CTT, nécessitait le calcul de scores individuels grâce à des imputations multiples des réponses manquantes au moyen d'un modèle d'ANOVA mixte à deux facteurs (un effet fixe correspondant à l'effet « items » et un effet aléatoire correspondant à l'effet « individus » – méthode « Two-Way »). En l'absence de DIF, ces deux méthodes fournissaient des résultats non biaisés même en cas de données manquantes informatives, c'est-à-dire dépendantes du trait latent que l'on cherchait à évaluer.

Il peut paraître surprenant que ces méthodes n'entraînent aucun biais en cas de données manquantes informatives, c'est-à-dire de données manquantes non aléatoires. Les méthodes habituellement recommandées en cas de données manquantes non aléatoires correspondent en effet à des méthodes de modélisation jointe de la variable d'intérêt et du processus de données manquantes [79], classiquement au moyen de modèles de sélection ou de modèles de types pattern-mixture [136]. Lors de l'utilisation des méthodes « Wald » ou « Two-Way », le processus responsable de données manquantes est ignoré et seule la variable d'intérêt est modélisée.

Une réflexion quant à la nature réelle de la typologie des données manquantes doit être menée en cas de données manquantes informatives et sur des PROs pour tenter d'expliquer ce phénomène.

Lorsque la probabilité d'observer une donnée manquante à un item donné était dépendante de la valeur du trait latent de l'individu interrogé, la donnée manquante était définie comme étant informative, puisque fonction d'une variable non observée correspondant à la variable d'intérêt.

On peut cependant considérer que pour un individu, l'ensemble des items d'un questionnaire constitue des évaluations répétées du même phénomène latent d'intérêt. Que l'on se place dans le cadre de la CTT ou de l'IRT, ces différentes évaluations sont supposées être indépendantes les unes des autres pour un individu donné : dans le cadre de la CTT, les erreurs associées aux réponses observées à deux items différents d'un même questionnaire ne doivent pas être corrélées entre elles (voir équation n°3 page 7) et dans le cadre de l'IRT, la propriété d'indépendance locale implique l'indépendance des réponses observées à chacun des items conditionnellement à la valeur du trait latent, *a fortiori* conditionnellement aux individus interrogés.

Si chacun des items utilisés est une mesure non biaisée du trait latent et si ces mesures sont toutes indépendantes entre elles, il doit être possible d'envisager l'évaluation du trait latent d'un individu dès lors que celui-ci a répondu à au moins l'un des items qui lui étaient soumis. Dit autrement, une

réponse observée à un item doit être une évaluation valide du trait latent d'un individu, que ce dernier ait ou non répondu aux autres items qui lui étaient soumis et quelle que soit la raison pour laquelle il n'y a pas répondu (c'est-à-dire que ces données manquantes soient informatives ou non). La répétition des mesures, par la constitution de questionnaires composés de différents items, doit alors permettre de fournir une mesure plus précise de ce trait latent.

Dans le cas de données manquantes informatives, la probabilité pour un individu donné d'observer une réponse manquante à un item dépend bien d'une variable non observée : son trait latent. Celui-ci doit pouvoir être appréhendé indépendamment par chacune des réponses aux items, qu'elles soient observées ou manquantes. Les seules réponses observées, fournissant chacune une évaluation non biaisée du trait latent, devraient donc permettre une appréhension valide du trait latent, indépendamment des réponses manquantes. Ainsi et en cas de réponse manquante informative, la probabilité d'observer une réponse manquante devrait pouvoir être expliquée soit par une variable non observée (le trait latent), soit par un ensemble de variables observées (l'ensemble des réponses observées aux items). Le seul cas pour lequel la probabilité de réponse manquante ne pourrait être expliquée que par une variable non observée correspondrait alors au cas où l'individu n'aurait répondu à aucun des items du questionnaire. Dit autrement, pourvu que l'individu ait répondu à au moins un item du questionnaire, les données manquantes informatives ne peuvent être considérées comme des données manquantes de type MNAR, puisque le phénomène latent responsable de l'absence de réponse peut être appréhendé par les seules réponses observées. Il doit donc être possible de proposer des méthodes d'analyse permettant d'obtenir des résultats non biaisés même en cas de données manquantes informatives, pourvu que les individus analysés aient répondu à au moins un des items qui leur étaient soumis. Des méthodes d'analyse basées sur le score peuvent être pertinentes pourvu que la méthode d'imputation des données manquantes – permettant le calcul du score – soit appropriée aux données analysées. La méthode « Two-Way » repose sur un processus d'imputations basé sur une modélisation des réponses des individus en fonction de leurs caractéristiques propres mais aussi de celles des items qui leur sont soumis. Les caractéristiques des individus et des items sont donc pris en compte pour imputer chaque réponse manquante, c'est-à-dire pour envisager ce qu'aurait pu être la réponse non observée d'un individu donné (dont les caractéristiques doivent pouvoir être estimées de façon valide à partir de ses seules réponses observées) à un item donné (dont les caractéristiques doivent tout autant pouvoir être estimées de façon valide à partir des seules réponses observées à ce dernier). Si une estimation valide des caractéristiques des individus et des items peut être envisageable uniquement à partir des réponses

observées au questionnaire, une imputation valide des réponses manquantes devrait aussi être possible dès lors qu'elle repose sur ces caractéristiques. Cette hypothèse devrait permettre de comprendre les propriétés remarquables de cette méthode « Two-Way » en présence de données manquantes, qu'elles soient informatives ou non informatives.

Les méthodes d'analyse basées sur l'IRT permettent quant à elles de tirer parti de l'intégralité de l'information disponible sans avoir à recourir à des procédés d'imputations, puisque celles-ci permettent l'analyse de chacune des réponses observées sans avoir à recourir à la notion de score.

Ainsi, une analyse de variables latentes peut être effectuée sur l'ensemble des individus ayant répondu à au moins l'un des items proposés si la méthode utilisée est appropriée (par exemple la méthode « Wald » ou la méthode « Two-Way »). Même si le processus responsable de l'apparition de données manquantes dépend du trait latent, on ne devrait pas considérer ces données manquantes comme étant de type MNAR, puisque ce trait latent peut être appréhendé par les seules réponses observées, indépendamment des réponses manquantes. Il paraît donc vraisemblable que des méthodes d'analyse appropriées, basées uniquement sur les réponses observées, puissent conduire à des résultats non biaisés même en cas de données manquantes informatives.

Ces méthodes d'analyse paraissent donc particulièrement intéressantes. La méthode « Two-Way » est d'interprétation aisée puisque celle-ci repose sur la notion de score, qui est d'usage courant dans le domaine des PROs. Les résultats présenteront alors un sens immédiat et concret pour la majorité des cliniciens et des chercheurs et seront facilement comparés soit aux résultats de la littérature, soit à des valeurs de score de référence en population générale, lorsque les questionnaires utilisés en disposent.

La méthode « Wald », quant à elle, permet effectivement de considérer le phénomène responsable du choix des réponses comme étant un phénomène latent et de l'analyser en tant que tel, plutôt que d'analyser une variable manifeste illustrative telle que pourrait être le score. Cette méthode est donc parfaitement congruente avec le concept de phénomène latent. Cependant, son interprétation reste délicate, tant par l'absence de valeurs de référence de traits latents en population générale que par la très grande confidentialité de son utilisation en recherche clinique, empêchant à l'heure actuelle toute possibilité de familiarité de la part de cliniciens ou de chercheurs avec celle-ci.

Il faut enfin remarquer que les méthodes « Wald » et « Two-Way » reposent toutes deux sur une

même hypothèse : les caractéristiques des items sont considérées identiques pour tous les individus quelles que soient leurs caractéristiques socio-culturelles, et constantes au cours du temps quelles que soient les spécificités de leur histoire personnelle.

Nous avons pu illustrer les possibles limites de cette hypothèse dans un cadre transversal en introduisant la notion de fonctionnement différentiel des items. Une stratégie possible de prise en compte de ce DIF a été proposée, permettant à la fois sa détection et sa prise en compte lors de l'analyse de bases de données. Cette stratégie n'est cependant pas la seule envisageable et d'autres méthodes pourraient être considérées. Ainsi, l'identification des items présentant un DIF pourrait être basée sur le critère d'Akaike, le critère d'information bayésien, ou des tests de rapports de vraisemblance par l'utilisation de processus ascendants, descendants ou itératifs mixtes. De même, la méthode proposée par Oort dans un cadre longitudinal [137] pourrait être adaptée au cas transversal pour le diagnostic du DIF. Enfin, la prise en compte d'une possible variation d'interprétation des items entre différents groupes d'individus et dans un cadre transversal (c'est-à-dire la prise en compte d'un possible DIF) impose de s'interroger sur la possibilité de variations d'interprétations au cours du temps, c'est-à-dire de façon longitudinale.

Ainsi, il est tout à fait vraisemblable que l'histoire individuelle, le chemin de vie ou l'évolution personnelle de chacun puisse constituer des facteurs externes de modification du langage, donc induire des variations d'interprétation et de compréhension de mêmes mots au cours du temps. Le langage ne peut donc être vu comme une construction figée pour chaque individu, mais doit au contraire être considéré comme un processus dynamique en perpétuel remodelage. Il est alors naturel de considérer possiblement variable au cours du temps l'interprétation que peuvent se faire des individus interrogés des items qui leur sont proposés, c'est-à-dire de possibles phénomènes de « response-shift » [137].

Il semble très artificiel de séparer les phénomènes de DIF et de « response-shift ». Ceux-ci correspondent à des phénomènes similaires considérés de deux points de vue différents : l'un transversal et l'autre longitudinal. Les méthodes développées pour l'identification et la prise en compte du DIF devront être adaptées à l'identification et la prise en compte de « response-shift » pour proposer une méthode globale d'identification et de prise en compte de variations d'interprétation des items lors de l'analyse de PROs, quel que soit le cadre dans lequel intervient cette variation d'interprétation.

Tout comme le DIF, le « response-shift » peut être modélisé à l'aide de modèles issus de la théorie

de la réponse à l'item, en intégrant cette fois dans un modèle longitudinal des interactions entre le temps et les difficultés des items. Il est alors évidemment possible de rechercher conjointement un phénomène de DIF et de « response-shift » lors d'une seule et même analyse, au moyen de méthodes basées sur la réalisation de modèles de complexité croissante. Si ces méthodes s'avèrent valides (c'est-à-dire non biaisées et de puissance satisfaisante), des solutions logicielles pourraient alors être proposées, permettant l'implémentation aisée de celles-ci. Le développement de telles solutions logicielles n'est cependant pas chose facile. Un logiciel d'analyse doit en effet non seulement permettre l'implémentation d'une méthode d'analyse donnée, mais aussi l'obtention de résultats interprétables, fournis dans des délais raisonnables. La complexité des modèles sur lesquels reposent les méthodes d'analyse telles qu'envisagées est telle qu'il est probable que les temps de calcul s'avèrent trop longs pour une utilisation de pratique courante. Ainsi, un programme additionnel Stata® a été développé permettant la mise en œuvre de la méthode «Wald-DIF » telle que décrite au chapitre 6.1.2 de ce mémoire. Ce programme repose sur l'emboîtement de modèles mixtes de types logistiques multinomiaux à risques adjacents de complexité croissante, dont le nombre de paramètres à estimer augmente avec le nombre d'items, de modalités de réponse et de covariables induisant un possible phénomène de DIF. Ce programme permet effectivement l'implémentation de la méthode « Wald-DIF », mais au prix de temps de calculs variables, de l'ordre de quelques minutes en l'absence de DIF à plusieurs jours si de nombreux items présentent un tel phénomène. Des solutions pourraient être envisagées pour permettre une accélération de ces temps de calcul, reposant non pas sur le développement de solutions logicielles basées sur l'utilisation de langages de programmation de haut niveau (tels que ceux utilisés par les logiciels de statistique tels que R ou Stata®), mais sur des programmes dédiés, réalisés dans des langages de programmation de plus bas niveau et possiblement compilés, tels que C++ ou Fortran. On pourra alors peut-être envisager de proposer des logiciels réellement utilisables en pratique quotidienne, car présentant des délais de calcul plus raisonnables.

8. CONCLUSION

L'étude de mesures subjectives est de plus en plus fréquente en recherche clinique. De telles mesures, permettant l'analyse de la subjectivité et du ressenti des patients, autorisent une appréhension différente de la notion de santé, des effets de maladies et de traitements. Accompagnées par l'étude de mesures objectives, elles permettent une compréhension plus globale

des processus physiopathologiques.

La réalisation de telles mesures subjectives est le plus souvent effectuée au moyen d'auto-questionnaires standardisés. Il peut alors arriver d'observer des réponses manquantes à ces questionnaires : un patient trop fatigué pourra s'arrêter dans le remplissage des items avant la fin du questionnaire ; un autre, jugeant l'une des questions trop sensible ou trop personnelle, pourra décider de ne pas y répondre.

Dans le cadre d'études cliniques, l'analyse de mesures subjectives passe habituellement par le calcul de scores. Lorsque les questionnaires comportent des données manquantes, ce calcul n'est plus directement possible. Différentes stratégies ont pu être proposées, les plus habituelles consistant à n'analyser que les individus ayant répondu à l'intégralité des items qui leur étaient soumis, ou à imputer les réponses manquantes de chaque individu par la moyenne de ses réponses observées. Nous avons pu montrer dans ce mémoire les limites de ces deux approches, à la fois biaisées et peu puissantes. Pourtant ces deux approches restent toujours actuellement les plus fréquemment utilisées, probablement en raison de leur facilité d'utilisation.

D'autres stratégies peuvent cependant être proposées. Dans ce mémoire, nous avons présenté deux autres approches : la méthode « Two-Way » et la méthode « Wald ». La méthode « Two-Way » permet le calcul de scores par la réalisation d'imputations multiples des réponses manquantes. Le processus d'imputation repose sur une modélisation des réponses aux items par un modèle mixte d'ANOVA à deux facteurs – l'un à effets fixes (l'effet « item »), l'autre à effets aléatoires (l'effet « individus »). La méthode « Wald », quant à elle, repose sur le test de la significativité d'une covariable incluse dans un modèle de la famille de Rasch à effets aléatoires et ne nécessite pas d'imputations des réponses manquantes. Un tel modèle permet la modélisation des réponses observées aux items là aussi par un paramètre associé aux individus (leur trait latent) et d'autres associés aux items (leurs difficultés), cette fois-ci au moyen d'une fonction de lien de type logistique. Ces deux méthodes présentent des propriétés à la fois très semblables et très intéressantes : une absence de biais et une puissance élevée même en cas de données manquantes informatives. Cependant, ces deux méthodes restent délicates à implémenter. Aussi, pour faciliter leur diffusion, nous avons développé des programmes additionnels au logiciel Stata® permettant leur réalisation.

Des questions restent actuellement en suspens, notamment concernant les problématiques liées à l'interprétation et à la compréhension par les personnes interrogées des items constituant les

questionnaires utilisés. Des méthodes d'analyse prenant en compte de possibles variations d'interprétation (c'est-à-dire des phénomènes de DIF ou de « response-shift ») sont envisageables, mais leurs propriétés méthodologiques sont actuellement inconnues. Il est maintenant nécessaire d'entreprendre des études de simulation pour déterminer leurs caractéristiques et ainsi identifier les méthodes d'analyse les plus pertinentes en cas de variations d'interprétation des items, tant dans un cadre transversal que longitudinal.

Bibliographie

- [1] Préambule à la Constitution de l'Organisation mondiale de la Santé, tel qu'adopté par la Conférence internationale sur la Santé, New York, 19-22 juin 1946 ; signé le 22 juillet 1946 par les représentants de 61 Etats. 1946 (Actes officiels de l'Organisation mondiale de la Santé, n°. 2, p. 100) et entré en vigueur le 7 avril 1948.
- [2] Dintiman G & Greenberg JS. *Health through discovery*. New York (Ed.). Random House, 1986.
- [3] Nordenfelt L. *Concepts and Measurement of Quality of Life in Health Care*. Dordrecht (Ed.). Kluwer Academic, 1994.
- [4] Circulaire ministérielle DH/PE1 n°10252 du 18 novembre 1992.
- [5] Jodelet D. Culture et pratiques de santé. *Nouvelle revue de psychosociologie* (2006) 1: pp. 219-239.
- [6] Jouet E, Flora L & Las Vergnas O. Construction et reconnaissance des savoirs expérimentiels des patients : note de synthèse. *Pratiques de formation - Analyses (n° spécial : "Usagers-experts : la part du savoir des malades dans le système de santé")* (2010) 58-59: pp. 13-94.
- [7] Le Breton D. *Anthropologie du corps et modernité*. PUF (Ed.). coll. « Quadrige Essais Débats », 2008.
- [8] Najman JM & Levine S. Evaluating the impact of medical care and technologies on the quality of life : a review and critique. *Social Science and Medicine* (1981) 15(2): pp. 107-115.
- [9] Björner JB, Fayers P & Idler H. Assessing quality of life in clinical trials: methods and practice. In *Oxford: oxford university press, 2005: 309–323*. Oxford (Ed.).2005.
- [10] Jacot W, Colinet B, Bertrand D, Lacombe S, Bozonnet MC, Daurès JP & Pujol JL. Quality of life and comorbidity score as prognostic determinants in non-small-cell lung cancer patients. *Ann Oncol* (2008) 19: pp. 1458-1464.
- [11] Park SH, Cho MS, Kim YS, Hong J, Nam E, Park J, Cho EK, Shin DB, Lee JH & Lee WK. Self-reported health-related quality of life predicts survival for patients with advanced gastric cancer treated with first-line chemotherapy. *Qual Life Res* (2008) 17: pp. 207-214.
- [12] Lee CK, Stockler MR, Coates AS, Gebiski V, Lord SJ, Simes RJ & Australian New Zealand Breast Cancer Trials Group. Self-reported health-related quality of life is an independent predictor of chemotherapy treatment benefit and toxicity in women with advanced breast cancer. *Br J Cancer* (2010) 102(9): pp. 1341-1347.
- [13] Svensson H, Hatschek T, Johansson H, Einbeigi Z & Brandberg Y. Health-related quality of life as prognostic factor for response, progression-free survival, and survival in women with metastatic breast cancer. *Med Oncol* (2012) 29(2): pp. 432-438.
- [14] Canghilhem G. *Le normal et le pathologique*. Paris (Ed.). PUF, Quadrige, 1966.
- [15] Leplège A & Coste J. *Mesure de la Santé Perceptuelle et de la Qualité de Vie : méthodes et applications*. Paris (Ed.). Estern, 2001.
- [16] De Bock E, Hardouin JB, Blanchin M, Le Neel T, Kubis G & Sébille V. Assessment of score-and Rasch-based methods for group comparison of longitudinal patient-reported outcomes with intermittent missing data (informative and non-informative). *Qual Life Res* (2014) Epub ahead of print.
- [17] Blanchin M, Hardouin JB, Le Neel T, Kubis G, Blanchard C, Mirallié E & Sébille V. Comparison of CTT and Rasch-based approaches for the analysis of longitudinal Patient Reported Outcomes. *Statistics in Medicine* (2011) 30(8): pp. 825-838.

- [18] Walters SJ, Campbell MJ & Lall R. Design and analysis of trials with quality of life as an outcome a practical guide. *Journal of Biopharmaceutical Statistics* (2001) 11: pp. 155-176.
- [19] Hambleton RK & Jones RW. Comparison of classical test theory and item response. *Educational Measurement Issues and Practice* (1993) 12(3): pp. 38-47.
- [20] Lord FM & Novick MR. Statistical theories of mental test scores. Reading M (Ed.). Addison-Wesley, 1968.
- [21] Magnusson D. Test theory. Reading M (Ed.). Addison-Wesley, 1967.
- [22] Gulliksen H. Theory of mental tests. New York (Ed.). Wiley, 1950.
- [23] Fan X. Item Response Theory and Classical Test Theory: An Empirical Comparison of their Item/Person Statistics. *Educational and Psychological Measurement* (1998) 58(3): pp. 357-381.
- [24] Rasch G. Probabilistic models for some intelligence and attainment tests. MESA Press, 1960.
- [25] Guttman L. **The basis for scalogram analysis**. In *Measurement and prediction. the american soldier vol. iv*. New York (Ed.). 1950.
- [26] Rasch G. On specific objectivity. An attempt at formalizing the request for generality and validity of scientific statements. *The Danish Yearbook of Philosophy* (1977) 14: pp. 58-94.
- [27] McKinley RL. Methods, plainly speaking: An introduction to Item Response Theory. *Measurement and Evaluation in Counseling and Development* (1989) 22: pp. 37-57.
- [28] Fischer GH & Molenaar IW. **Derivations of the Rasch model**. In *Rasch models, foundations, recent developments and applications*. Springer (Ed.). 1995. pp. 15-38.
- [29] Andersen EB. Sufficient statistics and latent trait models. *Psychometrika* (1977) 42(1): pp. 69-81.
- [30] Andersen EB. Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society, Series B* (1970) 32: pp. 283-301.
- [31] Fisher GH & Molenaar IW. Estimation of item parameters. In *Rasch models, foundations, recent developments and applications*. Springer (Ed.). 1995.
- [32] Pfanzagl J. On the consistency of conditionnal maximum likelihood estimators. *Annals of the Institute of Statistical Mathematics* (1993) 45(4): p. 703–719.
- [33] Hoijtink H & Boomsma A. On person parameter estimation in the dichotomous Rasch model. In *Rasch models, foundations, recent developments and applications*. G. H. Fisher & I. W. Molenaar (Ed.). 1995.
- [34] Zwinderman AH. Pairwise parameter estimation in rasch models. *Applied Psychological Measurement* (1995) 19(4): pp. 369-375.
- [35] Andrich D & Luo G. Conditional pairwise estimation in the rasch model for ordered response categories using principal components. *Journal of Applied Measurement* (2003) 4(3): pp. 205-221.
- [36] Lehmann EL & Casella G. Theory of Point estimation, 2nd Edition. Springer (Ed.). New York, 1998.
- [37] Cameron AC & Trivedi PK. Microeconometrics: Methods and Applications. Cambridge University Press (Ed.). New York, 2005.
- [38] Bock RD & Lieberman M. Fitting a response model for n dichotomously scored items. *Psychometrika* (1970) 35: pp. 189-197.
- [39] Mislevy RJ. Estimating latent distributions. *Psychometrika* (1984) 49(3): pp. 359-381.
- [40] Thissen D. Marginal maximum likelihood estimation for the one- parameter logistic model. *Psychometrika* (1982) 47: pp. 175-186.
- [41] Zwinderman AH. A generalized rasch model for manifest predictors. *Psychometrika* (1991) 56: pp. 589-600.
- [42] Christensen KB, Bjornera JB, Kreinerb S & Petersen JH. Latent Regression in Loglinear

- Rasch Models. *Communications in Statistics - Theory and Methods* (2004) 33(6): pp. 1295-1313.
- [43] Christensen KB. **Latent Covariates in Generalized Linear Models: A Rasch Model Approach**. In *Advances in statistical methods for the health sciences*. Auget (Ed.). 2007. pp. 95-108.
- [44] Hunt SM & Mc Ewen J. The development of a subjective health indicator. *Sociol Health Illn* (1980) 2(3): p. 231–246.
- [45] Hunt SM, McKenna SP, MCEwen J, Williams J & Papp E. The Nottingham Health Profile: subjective health status and medical consultations. *Social Science and Medicine* (1981) 15(3): pp. 221-229.
- [46] Leplège A, Ecosse E, Verdier A & Perneger T. The French SF-36 Health Survey: translation, cultural adaptation and preliminary psychometric evaluation. *J Clin Epidemiol* (1998) 51: pp. 1013-1023.
- [47] Ware JE & Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical Care* (1992) 30(6): pp. 473-483.
- [48] Andrich D. A rating formulation for ordered response categories. *Psychometrika* (1978) 43: p. 561–573.
- [49] Masters GN. A Rasch model for partial credit scoring. *Psychometrika* (1982) 47: pp. 149-174.
- [50] Agresti A. *Categorical Data Analysis*. Hoboken (Ed.). Wiley, 2002.
- [51] Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, Filiberti A, Flechtner H, Fleishman SB, de Haes JC & et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* (1993) 85(5): pp. 365-376.
- [52] Aaronson NK, Ahmedzai S, Bergman B, Bullinger M, Cull A, Duez NJ, Filiberti A, Flechtner H, Fleishman SB, de Haes JC, Kaasa S, Klee M, Osoba D, Razavi D, Rofe PB, Schraub S, Sneeuw K, Sullivan M & Takeda F. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst* (1993) 85: pp. 365-376.
- [53] Rabe-Hesketh S, Skrondal A & Pickles A. *GLLAMM Manual*. University of California-Berkeley, Division of Biostatistics, Working Paper Series, 2004.
- [54] Rubin DB. *Multiple imputation for nonresponse in surveys*. John Wiley & Sons, New Jersey, 1987.
- [55] Skrondal A & Rabe-Hesketh S. *Generalized latent variable modeling multilevel, longitudinal and structural equation models*. Boca Raton (Ed.). Chapman and Hall CRC, 2004.
- [56] Glas CAW, Geerlings H, van de Laar MAFJ & Taal E. Analysis of longitudinal randomized clinical trials using item response models. *Contemporary Clinical Trials* (2009) 30: pp. 158-170.
- [57] Wu M. The role of plausible values in large-scale surveys. *Studies in Educational Evaluation* (2005) 31: pp. 114-128.
- [58] Mislevy RJ, Johnson EG & Muraki E. Scaling procedures in NAEP. *Journal of Educational Statistics* (1992) 17: pp. 131-154.
- [59] Lord FM. A strong true-score theory, with applications. *Psychometrika* (1965) 30: pp. 239-270.
- [60] Lord FM. Estimating true-score distributions in psychological testing (an empirical bayes estimation problem). *Psychometrika* (1969) 34(3): pp. 259-299.
- [61] Mislevy RJ. Randomization-based inference about latent variables from complex samples. *Psychometrika* (1991) 56(2): pp. 177-196.
- [62] Mislevy RJ. Estimating population characteristics from sparse matrix samples of item

- responses. *Journal of Educational Measurement* (1992) 29(2): pp. 133-161.
- [63] Tang L, Song J, Belin TR & Unützer J. A comparison of imputation methods in a longitudinal randomized clinical trial. *Statistics in medicine* (2005) 24: pp. 2111-2128.
- [64] OECD (2012) PISA 2009 Technical Report, PISA, OECD Publishing.
<http://dx.doi.org/10.1787/9789264167872-en>
- [65] Allen NL, Donoghue JR & Schoeps TL (2001) The NAEP 1998 Technical Report. NCES Number: 2001509. <http://nces.ed.gov/pubsearch/pubsinfo.asp?pubid=2001509>
- [66] Yamamoto K & Kulick E. **Scaling Methodology and Procedures for the TIMSS Mathematics and Science Scales**. In *Timss 1999 technical report*. Chestnut Hill (Ed.). 2000.
- [67] Hardouin JB. Rasch analysis Estimation and tests with raschtest. *The Stata Journal* (2007) 7(1): pp. 22-44.
- [68] Holland PW & Hoskens M. Classical test theory as a first-order item response theory: application to true-score prediction from a possibly nonparallel test. *Pychometrika* (2003) 68(1): pp. 123-149.
- [69] Ross SM. **Computing Expectations by Conditioning**. In *Introduction to probability models, 10th edition*. Academic Press (Ed.). 2009. pp. 106-117.
- [70] Hamel JF, Hardouin JB, Le Neel T, Kubis G, Sébille V. Rasch models to detect a group effect for Patients Reported Outcomes. Communication affichée : 31st annual conference of the international society of clinical biostatistics (ISCB), August 31-September 2, 2010.
- [71] Hamel JF, Hardouin JB, Le Neel T, Kubis G, Roquelaure Y & Sebille V. Biases and Power for Groups Comparison on Subjective Health Measurements. *Plos One* (2012) 7(10): p. e44695.
- [72] Cohen J. *Statistical Power Analysis for the Behavioral Sciences* (second ed.). Lawrence Erlbaum Associates, 1988.
- [73] StataCorp. 2013. *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP. 11.
- [74] Rabe-Hesketh S, Pickles A & Taylor C. Generalised, linear, latent and mixed models. *Stata Technical Bulletin* (2000) 53: pp. 47-57.
- [75] Eggen TJHM. On the loss of information in conditional maximum likelihood estimation of item parameters. *Psychometrika* (2000) 65(3): pp. 337-362.
- [76] Sébille V, Hardouin JB, Le Néel T, Kubis G, Boyer F, Guillemin F & Falissard B. Methodological issues regarding power of classical test theory (CTT) and item response theory (IRT)-based approaches for the comparison of patient-reported outcomes in two groups of patients - a simulated study. *BMC Medical Research Methodology* (2010) 25: pp. 10-24.
- [77] Matsuda T, Marche H, Grosclaude P & Clement S. Participation behavior of bladder cancer survivors in a medical follow-up survey on quality of life in France. *European Journal of Epidemiology* (2004) 19: pp. 313-321.
- [78] Hensing TA, Peterman AH, Schell MJ, Lee JH & Socinski MA. The Impact of Age on Toxicity, Response Rate, Quality of Life, and Survival in Patients with Advanced, Stage IIIB or IV Non-small Cell Lung Carcinoma Treated with Carboplatin and Paclitaxel. *American Cancer Society* (2003) 98(4): pp. 779-788.
- [79] Rubin DB. Inference and missing data. *Biometrika* (1976) 63: pp. 581-592.
- [80] Little RJA & Rubin DB. *Statistical Analysis with Missing Data*. Wiley & Sons (Ed.). New York, 1987.
- [81] Schafer JL & Graham JW. Missing Data: Our View of the State of the Art. *Psychological Methods* (2002) 7(2): pp. 147-177.
- [82] Ibrahim JG, Chen MH, Lipsitz SR & Herring A. Missing-Data Methods for Generalized Linear Models: A Comparative Review. *Journal of the American Statistical Association* (2005) 100(469): pp. 332-346.

- [83] Little RJA. Modeling the Drop-Out Mechanism in Longitudinal Studies. *Journal of the American Statistical Association* (1995) 90: pp. 1112-1121.
- [84] Diggle P & Kenward MG. Informative drop-out in longitudinal data analysis. *Applied Statistics* (1994) 43(1): pp. 49-93.
- [85] Little RJA. Discussion to Diggle, P.J. and Kenward, M.G. Informative dropout in longitudinal data analysis. *Applied Statistics* (1994b) 43: p. 78.
- [86] Muller-Nordhorn J, Roll S & Willich SN. Comparison of the short form (SF)-12 health status instrument with the SF-36 in patients with coronary heart disease. *Heart* (2004) 90: p. 523–527.
- [87] Peyre H, Coste J & Leplège A. Identifying type and determinants of missing items in quality of life questionnaires: Application to the SF-36 French version of the 2003 Decennial Health Survey. *Health and Quality of Life Outcomes* (2010) 8:16: on line.
- [88] Hopwood P, Stephens RJ & Machin D. Approaches to the analysis of quality of life data: experiences gained from a Medical Research Council Lung Cancer Working Party palliative chemotherapy trial. *Quality of Life Research* (1994) 3: pp. 339-352.
- [89] Hamel JF, Hardouin JB, Le Neel T, Kubis G, Sébille V. Rasch models to detect a group effect for Patients Reported Outcomes with possibly informative missing data. Communication orale : International Workshop on "Patient Reported Outcomes and Quality of Life", Annales de l'ISUP (2011), 55(2): 49-52, 2011.
- [90] Hamel JF, Hardouin JB, Le Neel T, Kubis G, Sébille V. Detection of a group effect for patients reported outcomes with Rasch models in the presence of possibly informative missing values. Communication orale : Congrès PROQOL: Patient Reported Outcomes and Quality of Life, 2011, Paris, France, 2011.
- [91] Hardouin JB, De Bock E, Blanchin M, Hamel JF, Sébille V. Interest of Item Response Theory (IRT) models for handling missing data in cross-sectional or longitudinal designs. Communication orale : Quels apports des mathématiques à la mesure de la qualité de vie en oncologie, 11-12 septembre 2012, Marseille, France, 2013.
- [92] Hamel JF, Sebille V, Le Neel T, Kubis G, Boyer FC & Hardouin JB. Study of different methods for comparing groups of patients by analysing patients reported outcomes when data is missing. *Statistical Methods in Medical Research* (Under review).
- [93] Van Ginkel JR, Van der Ark LA, Sijtsma K & Vermunt JK. Two-way imputation: A Bayesian method for estimating missing scores in tests and questionnaires, and an accurate approximation. *Computational Statistics & Data Analysis* (2007) 51: pp. 4013-4027.
- [94] Bernaards CA & Sijtsma K. Influence of imputation and EM methods on factor analysis when item nonresponse in questionnaire data is nonignorable. *Multivariate Behav. Res.* (2000) 35: pp. 321-364.
- [95] Brennan RL. Generalizability Theory. Springer (Ed.). New York, 2001.
- [96] Tanner MA & Wong WH. The calculation of posterior distributions by data augmentation. *J. Am. Statist. Assoc* (1987) 82: pp. 528-540.
- [97] Dondersa ART, Van der Heijdenc GJMG, Stijnen T & Moons KGM. *Review: A gentle introduction to imputation of missing values* (2006) *Journal of Clinical Epidemiology*: pp. 1087-1091.
- [98] Hardouin JB, Conroy R & Sébille V. Imputation by the mean score should be avoided when validating a Patient Reported Outcomes questionnaire by a Rasch model in presence of informative missing data. *BMC Med Res Methodol.* (2011) 11: p. 115.
- [99] White IR & Carlin JB. Bias and efficiency of multiple imputation compared with complete-case analysis for missing covariate values. *Statistics in Medicine* (2010) 29(28): pp. 2920-2931.

- [100] Holland PW & Wainer H. Differential item functioning. Hillsdale N (Ed.). Lawrence Erlbaum, 1993.
- [101] Perrot B, Le Neel T, Hamel JF, Blanchin M, Sébille V, Hardouin JB. Online analyses of PRO using CTT and IRT methods for the validation or the analysis of scale or for the planning of studies. Communication affichée : 21th annual conference of ISOQOL, 15-18 october 2014, Berlin, Germany, 2014
- [102] Perrot B, Le Neel T, Hamel JF, Blanchin M, Sébille V, Hardouin JB. Online analyses of PRO using CTT and IRT methods for the validation or the analysis of scales or for the planning of studies. Communication affichée : Chronic diseases, adaptation and perceived health: stakes and future - Young Researchers Days, 3-6 juin 2014, Nancy, 2014.
- [103] Perrot B, Le Neel T, Hamel JF, Blanchin M, Sébille V, Hardouin JB. Online analyses of PRO using CTT and IRT methods for the validation or the analysis of scale or for the planning of studies. Communication affichée : 20th annual conference of the ISOQOL, October 9-12, 2013, Miami (USA), 2013.
- [104] PCMODEL: Stata module to estimate parameters of a Partial Credit Model or a Rating Scale Model by MML. <http://ideas.repec.org/c/boc/bocode/s457788.html>.
- [105] PCMTEST: Stata module to test the fit of a Partial Credit Model or a Rating Scale Model estimated using pmodel. <http://ideas.repec.org/c/boc/bocode/s457789.html>.
- [106] Glas CAW & Verhelst ND. **Tests of Fit for Polytomous Rasch Models**. In *Rasch models: foundations, recent developments, and applications*. G. H. Fischer & I. W. Molenaar (Ed.). 1995. pp. 325-353.
- [107] Glas CAW & Verhelst ND. **Testing the Rasch model**. In *Rasch models: foundations, recent developments, and applications*. G. H. Fischer & I. W. Molenaar (Ed.). 1995. pp. 69-95.
- [108] Glas CAW. The derivation of some tests for the Rasch model from the multinomial distribution. *Psychometrika* (1988) 53: pp. 525-546.
- [109] McFadden D. **Quantitative methods for analysing travel behaviour of individuals: some recent developments**. In *Behavioural travel modelling*. Hensher, D. A & Stopher, P.R. (Ed.). 1979. pp. 279-318.
- [110] Self SG & Mauritsen RH. Power and sample size calculations for generalized linear models. *Biometrics* (1988) 44: pp. 79-86.
- [111] Langer E. The illusion of control. *Journal of personality and social psychology* (1975) 31: pp. 311-328.
- [112] Ladouceur R, Sylvain C & Boutin C. Le jeu pathologique. *Revue québécoise de psychologie* (2000) 21 (1): p. 21–35.
- [113] Dannewitz H & Weatherly J. Investigating the illusion of control in mildly depressed and non depressed individuals during video-poker play. *The journal of psychology* (2007) 141(3): p. 307–319.
- [114] Moodie C. Student gambling, erroneous cognitions and awareness of treatment in Scotland. *Journal of gambling issues* (2008) 21: pp. 30-54.
- [115] Bonnaire C, Lejoyeux M & Dardennes M. Sensation seeking in a french population of pathological gamblers: comparison with regular and nongambler. *Psychological reports* (2004) 94 (3): p. 1361–1371.
- [116] Miller N & Currie S. A canadian population level analysis of the role of irrational gambling cognitions and risky gambling practices as correlates of gambling intensity and pathological gambling. *Journal of gambling studies* (2008) 24: p. 257–274.
- [117] Breen R & Zuckerman M. Chasing in gambling behavior: personality and cognitive determinants. *Personality and individual differences* (1999) 27: p. 1097–1111.
- [118] Bouju G, Hardouin JB, Boutin C, Gorwood P, Le Bourvellec JD, Feuillet F, Venisse JL &

Grall-Bronnec M. A shorter and multidimensional version of the Gambling Attitudes and Beliefs Survey (GABS-23). *J Gambl Stud* (2014) 30(2): pp. 349-367.

- [119] miceadds: Some additional multiple imputation functions, especially for mice. <http://cran.r-project.org/web/packages/miceadds/index.html>.
- [120] Stata Multiple Imputation Reference Manual. <http://www.stata.com/manuals13/mi.pdf>.
- [121] MI_TWOWAY: Stata module for computing scores on questionnaires containing missing item responses. <http://ideas.repec.org/c/boc/bocode/s457901.html>.
- [122] Gelman A, Carlin JB, Stern HS & Rubin DB. Bayesian Data Analysis, second ed. Chapman & Hall (Ed.). London, 2003.
- [123] Sijtsma K & Van der Ark LA. Investigation and treatment of missing item scores in test and questionnaire data. *Multivariate Behavioral Research* (2003) 38: pp. 505-528.
- [124] Labov W. The Social Stratification of English in New York City. Cambridge (Ed.). Cambridge University Press, 1966.
- [125] Woolard K. Language variation and cultural hegemony: Towards an integration of sociolinguistic and social theory. *American ethnologist* (1985) 12: p. 738–748.
- [126] Lakoff R. Language and woman's place. New York (Ed.). Harper & Row, 1975.
- [127] Eckert P & McConnell-Ginet S. Language and gender. Cambridge (Ed.). Cambridge University Press, 2003.
- [128] Hamel JF, Hardouin JB, Roquelaure Y, Sébille V. Assessing quality of life of patients suffering from chronic low back pain in the presence of DIF. Communication orale : Chronic diseases, adaptation and perceived health: stakes and future - Young Researchers Days, 3-6 juin 2014, Nancy, 2014.
- [129] Smith LL & Reise SP. Gender differences on negative affectivity: an IRT study of differential item functioning on the Multidimensional Personality Questionnaire Stress Reaction Scale. *J Pers Soc Psychol* (1998) 75(5): pp. 1350-1362.
- [130] Gao W, Stark D, Bennett MI, Siegert RJ, Murray S & Higginson IJ. Using the 12-item General Health Questionnaire to screen psychological distress from survivorship to end-of-life care: dimensionality and item quality. *Psychooncology* (2012) 21(9): pp. 954-961.
- [131] Sharp C, Michonski J, Steinberg L, Fowler JC, Frueh BC & Oldham JM. An investigation of differential item functioning across gender of BPD criteria. *J Abnorm Psychol* (2014) 123(1): pp. 231-236.
- [132] Ware JE, Kosinski M, Gandek B, Aaronson NK, Apolone G, Bech P, Brazier J, Bullinger M, Kaasa S, Leplège A, Prieto L & Sullivan M. The factor structure of the SF-36 Health Survey in 10 countries: results from the IQOLA Project. International Quality of Life Assessment. *J Clin Epidemiol.* (1998) 51(11): pp. 1159-1165.
- [133] Harding S. Whose science? Whose knowledge? Thinking from women's lives. Ithaca N (Ed.). Cornell University Press, 1991.
- [134] Kramarae C. Women and men speaking. Rowley M (Ed.). Newbury House, 1981.
- [135] Tannen D. You just don't understand. New York (Ed.). Ballantine, 1990.
- [136] Heckman J. The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. *Annals of Economic and Social Measurement* (1976) 5: p. 475–492.
- [137] Oort FJ, Visser MR & Sprangers MA. An application of structural equation modeling to detect response shifts and true change in quality of life data from cancer patients undergoing invasive surgery. *Qual Life Res* (2005) 14(3): pp. 599-609.

9. ANNEXES

<i>Diff.</i>	D_{Diff}	j	n	Δ	Δ_s	σ^2_s	CTT		IRT										
							Test t	CML				MML				MML-Cov			
								WML	PV	MI	EAP	PV	MI	² steps	EAP	PV	MI	Wald	
			800	0	0.000	4.897	-0.004	-0.002	0.002	0.000	-0.001	0.004	0.001	-0.002	-0.002	-0.007	-0.003	-0.002	
				0.2	0.353	4.890	0.346	0.199	0.202	0.201	0.130	0.136	0.132	0.201	0.201	0.196	0.200	0.203	
				0.5	0.880	4.851	0.883	<i>0.508</i>	<i>0.511</i>	<i>0.510</i>	0.339	0.345	0.341	<i>0.514</i>	<i>0.514</i>	<i>0.509</i>	<i>0.513</i>	0.518	
				0.8	1.404	4.780	1.411	<i>0.812</i>	<i>0.816</i>	<i>0.814</i>	0.561	0.566	0.563	<i>0.825</i>	<i>0.825</i>	<i>0.820</i>	<i>0.823</i>	0.830	

Tableau 29: Différences moyennes estimées entre les deux groupes comparés par les différentes méthodologies de comparaison de groupes par analyse de mesures subjectives, en fonction des différentes combinaisons de paramètres de simulation (différence de scores moyens pour la méthode CTT_{score} , différence de traits latents moyens pour les méthodes IRT)

Est. : Difficultés estimées lors de l'analyse, *Connu* : difficultés considérées parfaitement connues lors de l'analyse, *Imp.* : difficultés considérées connues de façon imprécise lors de l'analyse, *Norm.* : distribution normale, *Bimo.* : mélange bimodal de distributions normales.

Les biais de position statistiquement significatifs sont indiqués par une différence moyenne reportée en italique, alors que les biais de position considérés importants sont indiqués par une différence moyenne reportée en gras.

Diff.	D_{Diff}	j	n	Δ	Δ_s	σ_s^2	Test t	CTT				IRT							
								CML				MML				MML-Cov			
								WML	PV	MI	EAP	PV	MI	2 steps	EAP	PV	MI	Wald	
200	0	0.000	4.897	4.899	1.643	2.315	3.130	0.702	1.075	1.552	1.072	0.698	1.095	1.581	1.069				
	0.2	0.353	4.890	4.876	1.636	2.301	3.130	0.701	1.073	1.553	1.066	0.692	1.084	1.570	1.062				
	0.5	0.880	4.851	4.839	1.634	2.317	3.138	0.723	1.109	1.594	1.066	0.689	1.084	1.572	1.061				
	0.8	1.404	4.780	4.794	1.647	2.345	3.188	0.774	1.176	1.685	1.079	0.697	1.093	1.596	1.073				
400	0	0.000	4.897	4.901	1.664	2.340	3.064	0.712	1.093	1.593	1.092	0.711	1.099	1.553	1.092				
	0.2	0.353	4.890	4.890	1.665	2.349	3.067	0.716	1.096	1.600	1.091	0.709	1.098	1.552	1.091				
	0.5	0.880	4.851	4.850	1.665	2.356	3.082	0.739	1.130	1.643	1.090	0.707	1.100	1.554	1.089				
	0.8	1.404	4.780	4.781	1.667	2.363	3.114	0.780	1.182	1.716	1.092	0.706	1.104	1.559	1.089				
800	0	0.000	4.897	4.893	1.636	2.311	3.163	0.691	1.068	1.518	1.066	0.690	1.065	1.580	1.069				
	0.2	0.353	4.890	4.889	1.637	2.313	3.173	0.696	1.073	1.527	1.068	0.691	1.069	1.584	1.070				
	0.5	0.880	4.851	4.853	1.641	2.331	3.195	0.722	1.107	1.572	1.071	0.691	1.068	1.590	1.072				
	0.8	1.404	4.780	4.785	1.640	2.334	3.215	0.764	1.165	1.647	1.073	0.690	1.075	1.593	1.072				

Tableau 30: Variances estimées des scores (pour la méthode CTT_{score}) et des traits latents (pour les méthodes IRT) par les différentes méthodologies de comparaison de groupes par analyse de mesures subjectives, en fonction des différentes combinaisons de paramètres de simulation.

Est. : Difficultés estimées lors de l'analyse, *Connu* : difficultés considérées parfaitement connues lors de l'analyse, *Imp.* : difficultés considérées connues de façon imprécise lors de l'analyse, *Norm.* : distribution normale, *Bimo.* : mélange bimodal de distributions normales.

Les biais de dispersion statistiquement significatifs sont indiqués par une variance reportée en italique et les biais de dispersion importants en pratique sont indiqués par une variance reportée en gras,

