

UNIVERSITE DE NANTES
UFR DE SCIENCES PHARMACEUTIQUES

ECOLE DOCTORALE BIOLOGIE SANTE

Année 2012

N ° attribué par la bibliothèque

**Validation, Planification et Analyse de
concepts latents en recherche clinique**

HABILITATION A DIRIGER LES RECHERCHES

Jean-Benoit HARDOUIN

Soutenance le 3 juillet 2012, devant le jury ci-dessous,

Rapporteur > Dr Franck BONNETAIN, Centre Georges-François Leclerc (CLC), Dijon

Rapporteur > Pr Joël COSTE, PU-PH, Université René Descartes - Paris V

Rapporteur > Pr Bruno FALISSARD, DR, INSERM U669, Paris

Examineur > Pr Francis GUILLEMIN, PU-PH, Université Nancy

Examineur > Pr Alain LEPLÈGE, Professeur des Universités, Université Paris VII

Examineur > Pr Véronique SÉBILLE, PU-PH, Université de Nantes

Examineur > Pr Florence TUBACH, PU-PH, Université Paris Diderot, INSERM U738, Paris

*"En théorie, il n'y a pas
de différence entre la théorie
et la pratique.
En pratique, il y en a."*

Table des matières

Remerciements	vii
1 Curriculum Vitae	1
Renseignements administratifs	1
Cursus Universitaire	1
Parcours Professionnel	2
Production scientifique	2
Articles dans des revues internationales	2
Articles dans des revues nationales et thèse de doctorat	4
Chapitres d’ouvrage	4
Articles en révision	4
Articles soumis ou en cours d’écriture	4
Communications internationales	6
Posters dans des conférences internationales	7
Communications nationales	8
Posters dans des conférences nationales	9
Macro-programmes sous SAS	10
Modules Stata	11
Rapports d’études	11
Encadrement doctoral et scientifique	13
Encadrement doctoral	13
Accompagnement scientifique de doctorant (sans statut d’encadrant)	14
Encadrement de Master 2 avec production scientifique	14
Autres encadrements d’étudiants avec production scientifique	14
Autres encadrements prévus ou n’ayant pas encore donné lieu à une valorisation scientifique	15
Rayonnement scientifique	15
Conférences invitées	15
Co-responsabilité de contrats publics ou privés	15
Participation à des contrats comme méthodologiste/statisticien	15
Validation scientifique (Reviewing)	17
Organisation de manifestations scientifiques	17
Organisation de formations	17
Responsabilités pédagogiques	17
Ecole doctorale Biologie Santé ED 502(Université de Nantes)	17

Master 2 Recherche Interrégional "Modélisation en Pharmacologie Clinique et Epidémiologie - MPCE" (Universités de Rennes 1, Nantes, Angers, Brest, Tours, Poitiers)	18
Master 1 "Bioinformatique et Biostatistique" (Université de Nantes)	18
UFR de Pharmacie de l'Université de Nantes	18
UFR de Médecine de l'Université de Nantes	18
Ecole Nationale de la Statistique et de l'Analyse de l'Information (ENSAI)	18
Responsabilités administratives	18
2 Bilan (1999-2011)	19
2.1 Parcours universitaire	19
2.2 Validation d'échelle	20
2.2.1 Travaux méthodologiques	21
Doctorat (1999-2005)	21
Impact de l'imputation de données manquantes en validation de questionnaire (2007-2011)	25
2.2.2 Validation d'échelles	25
Distorsions cognitives des joueurs pathologiques	25
Mesure du raisonnement médical	26
Mesure de la qualité de vie des patients atteints de malformations anorectales	27
Mesure de la gravité de la Pharmacodépendance (2008-2012)	28
Construction d'un questionnaire de qualité de vie pour les patients atteints de maladies neuromusculaires (2007-2012)	28
2.3 Planification d'études portant sur des mesures subjectives	28
2.3.1 Comparaison des approches CTT et IRT : de la comparaison de groupes au projet PLANIPRO (depuis 2007)	28
2.3.2 Nombre de sujets nécessaires pour la validation de questionnaire (Depuis 2011)	30
2.4 Analyse de données subjectives	31
2.4.1 Travaux méthodologiques	31
Analyses séquentielles sur données subjectives (2004-2008)	31
Comparaison des méthodes CTT et IRT pour l'analyse de données transversales (depuis 2010)	32
Comparaison des méthodes CTT et IRT pour l'analyse de données longitudinales (depuis 2008)	33
Estimation des paramètres du modèle de Rasch (2009-2011)	34
L'analyse en classes latentes (Depuis 2009)	34
2.4.2 Travaux appliqués	35
Analyse longitudinale de la qualité de vie et du coping des patientes atteintes d'un cancer du sein et de leur accompagnant (2007-2009)	35
Analyse spatio temporelle de la qualité de vie en population générale (Depuis 2009)	35
Education Médicale (Depuis 2007)	36
Pharmacodépendance (Depuis 2007)	37
Analyse du vécu des étudiants en Première Année Commune des Etudes de Santé à Nantes (2011)	37
2.5 Recherche Clinique	38

2.6	Epidémiologie et Santé Publique	39
2.7	Conclusion	39
2.7.1	Le modèle de Rasch présente-t-il un intérêt ?	39
2.7.2	Et les autres modèles de l'IRT	41
2.7.3	Les études de simulations de Monte Carlo	42
3	Perspectives et Projets	45
3.1	Détermination du nombre de sujets nécessaires pour l'analyse de mesures subjectives	45
3.1.1	Le projet ANR PLANIPRO (2011-2013)	45
3.1.2	L'interprétation de la différence sur le trait latent entre deux groupes de patients, pour définir la Différence Minimale Cliniquement Pertinente	46
3.1.3	Le nombre de sujets nécessaires pour la validation d'échelles	48
3.2	Validation d'échelles	48
3.2.1	Validation de la version française de l'échelle Impact of Cancer v2	48
3.2.2	Validation de la version française de l'Index de la Fonction Sexuelle Féminine (IFSF)	49
3.2.3	Validation du questionnaire de qualité de vie spécifiques aux maladies neuromusculaires	49
3.3	Comparaison de l'analyse des mesures subjectives entre la CTT et l'IRT	49
3.3.1	Etudes transversales	49
3.3.2	Etudes longitudinales	51
3.4	Autres perspectives à moyen terme	51
3.4.1	Création d'un outil de caractérisation des pharmacodépendances médicamenteuses	51
3.4.2	Recherche Clinique	52
3.4.3	Enseignement : le parcours M1 Bioinformatique-Biostatistique du Master Biologie Santé	52
4	Cinq publications représentatives du travail de recherche	55
4.1	Towards power and sample size calculations for the comparison of two groups of patients with Item Response Theory models	55
4.2	Nonparametric Item Response Theory using Stata	67
4.3	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 simulation study	91
4.4	Comparison of CTT and Rasch-based approaches for the analysis of longitudinal patient reported outcomes	103
4.5	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	119
5	Fiche de synthèse des encadrements et publications associées	135
5.1	Encadrement de thèse	135
5.2	Accompagnement scientifique de doctorant (sans statut d'encadrant)	136
5.3	Encadrement de Master 2 avec production scientifique	136
5.4	Autres encadrements d'étudiants avec production scientifique	137
5.5	Autres encadrements prévus ou n'ayant pas encore donné lieu à une valorisation scientifique	137

Remerciements

Mes remerciements vont tout d'abord à toi Véronique, pour m'avoir donné ma chance à Nantes, m'avoir soutenu, et associé à tes projets et en particulier dans l'aventure un peu folle à l'époque de l'EA 4275. Que de chemin parcouru depuis près de 6 ans, grâce à ton dynamisme et ton obstination ! J'ai eu beaucoup de chance de croiser ton chemin, et je sais que je n'en serai pas là sans ton soutien, ta bonne humeur et ta rigueur. Merci pour toute les fois où tu as corrigé mon anglais défectueux, j'essaye de continuer de progresser....

Merci aux membres de mon jury, et en particulier à mes rapporteurs, d'avoir relu ce travail, pour vos remarques, vos commentaires, qui, j'en suis certain, m'aideront à continuer d'avancer dans ma recherche.

Merci aussi à tous mes collaborateurs proches dans mes travaux de recherche sur Nantes, que ce soit les plus "anciens" comme Myriam, Jeff, Fanny, Elodie, Marie, Gaëlle, Pierre, Bernard, Caroline, Lise, Morgane, Angélique, Philippe, Yohann ou Etienne, ou d'autres plus récents comme Leila, Emmanuelle, Alexandra ou Mohand. Que de joies à collaborer avec vous tous, souvent dans la bonne humeur, parfois plus dans le stress et l'urgence, mais toujours dans le respect les uns des autres. J'ai beaucoup de chance de travailler avec vous tous !

J'ai aussi une pensée pour les stagiaires que j'ai eu la joie d'encadrer ou de coencadrer, Julie, Sarah, Laura, Jean-Damien, Pauline, Estelle, Mathilde, Emilie, Gaëlle. Vos personnalités toutes différentes, vos façons de travailler, vos connaissances m'ont aussi fait progresser. J'espère avoir pu vous transmettre un peu de mon expérience.

Merci aussi aux autres membres de l'EA4275, avec qui je travaille moins au jour le jour. Votre rigueur et vos compétences m'aident aussi à avancer et à découvrir chaque jour de nouveaux horizons dans le savoir humain.

J'ai une petite pensée pour toi aussi Mounir. Nos chemins de recherche ont certes bifurqué dans la forme, mais merci pour toutes les chances que tu m'as donné, les congrès dans lesquels tu m'as envoyé qui ont toujours été source de richesse, les personnes que tu m'as fait découvrir et ton soutien durant ma thèse.

Merci aussi à Lucie, Christelle, Cécile et Cécile. J'ai apprécié nos collaborations sur les STIC ou autres projets hospitaliers, toujours enthousiasmantes, et enrichissantes.

Merci aux investigateurs qui m'ont fait confiance comme méthodologiste et/ou statisticien.

Merci à mes parents qui m'ont appris à ne pas me satisfaire du minimum (et ce n'était pas gagné).

Merci à toi ma Roudoune pour ton amour, ta patience, tes certitudes, ton opiniâtreté, ta présence. Tu m'aides chaque jour à aller de l'avant et à me dépasser un peu plus.

Et puis un petit message personnel à mes 4 loulous :

- Orane (Louloute), ne te contente pas du minimum, va de l'avant, tu as un caractère suffisant pour pouvoir enfoncer les portes et faire de belles choses, profite-en, et n'oublie pas de toujours exploiter tes dons (devine lesquels!) et de faire ressortir ce qui est beau en toi!
- Théophile (Théoph'), continue sur ta lancée, ne flanche pas, reste exigeant avec toi même, serviable et curieux. Reste rigoureux, mais fais juste attention à ne pas t'enfermer dans un perfectionnisme trop pointilleux qui nuirait à ta créativité!
- Philémon (Phiphi), tu as des qualités pour aller loin. Apprends à ne pas te laisser envahir par tes sentiments, pour rester concentré sur ce que tu fais et sur la direction que tu souhaites suivre!
- Isaline (Poupette), restes bonne vivante, souriante et rigolote. On t'aime comme cela!

Chapitre 1

Curriculum Vitae

RENSEIGNEMENTS ADMINISTRATIFS

JEAN-BENOIT HARDOUIN

Adresse personnelle :

16 rue du pont de l'angle
44160 BESNE

Adresse professionnelle :

EA4275 "Biostatistique, Pharmacoépidémiologie et Mesure Subjective en Santé"
UFR de Médecine et des Sciences pharmaceutiques
1 rue Gaston Veil
44000 Nantes

Date et lieu de naissance :

3 avril 1976 à Yvetot (76)

Situation familiale :

Marié, 4 enfants

CURSUS UNIVERSITAIRE

- **2005** : Thèse de doctorat "Construction d'échelles d'items unidimensionnelles en Qualité de Vie" en "*Recherche Clinique, Innovations Technologiques et Santé Publique, spécialité Biostatistique*" - Université René Descartes - Paris 5, mention très honorable
- **1999** : Diplôme d'Etude Approfondies (DEA) de *Santé Publique* - Université René Descartes - Paris 5, mention Bien
- **1998** : Maîtrise Professionnelle "*Génie Informatique et Statistique*" (*GIS*) - Université de Bretagne Sud, mention Assez Bien
- **1996** : DUT "*Statistique et Traitement Informatique des Données*" (*STID*) - Université Lille 2
- **1994** : Baccalauréat C, mention Assez bien

PARCOURS PROFESSIONNEL

- **Depuis janvier 2010** : Ingénieur Hospitalier en méthodologie/Biostatistique à la plateforme de Biométrie du CHU de Nantes
- **Depuis octobre 2007** : Maître de conférences en Biostatistique à l'UFR de Pharmacie de l'université de Nantes
- **2006-2007** : ATER en Biostatistique à l'UFR de Pharmacie de l'université de Nantes
- **2001-2006** : Chargé de missions à l'Observatoire Régional de la Santé (ORS) de la région Centre (Orléans)
- **1999-2001** : Allocataire de recherche de l'université René Descartes - Paris 5

PRODUCTION SCIENTIFIQUE

Articles dans des revues internationales

- [1] M. Blanchin, J.-B. Hardouin, T. Le Néel, G. Kubis, C. Blanchard, E. Mirallié, and V. Sébille. Comparison of CTT and Rasch-based approaches for the analysis of longitudinal Patient Reported Outcomes. *Statistics in Medicine* [IF=2.33], 30(8) :825–838, April 2011. PMID : 21432877.
- [2] M. Blanchin, J.-B. Hardouin, T. Le Néel, G. Kubis, and V. Sébille. Analysis of longitudinal Patient-Reported Outcomes with informative and non-informative dropout : Comparison of CTT and Rasch-based methods. *International Journal of Applied Mathematics & Statistics (IJAMAS)* [IF inconnu], 24(SI-11A), 2011.
- [3] A. Bonnaud-Antignac, J.-B. Hardouin, J. Léger, A. Hartmann, F. Dravet, and V. Sébille. Quality of life and coping of women treated for breast cancer and their caregiver. What are the interactions? *Accepté pour publication dans Journal of Clinical Psychology in Medical Settings* [IF=1.506], 2012.
- [4] M. Bronnec, G. Bouju, J.-L. Vénisse, J.-D. Le Bourvellec, and J.-B. Hardouin. Psychometrics validation of the french version of the Gambling Related Cognition Scale (GRCS). *Accepté pour publication dans Journal of Gambling Issues* [IF=1.303], 2012.
- [5] M.-F. de La Cochetière, E. Montassier, J.-B. Hardouin, T. Carton, F. Le Vacon, T. Durand, V. Lalande, J.-C. Petit, G. Potel, and L. Beaugerie. Human intestinal microbiota gene risk factors for antibiotic-associated diarrhea : perspectives for prevention. *Microbial Ecology* [IF=2.88], 59(4) :830–837, May 2010. PMID : 20186408.
- [6] M. -L. Feddag, J. -B. Hardouin, and V. Sebille. Pairwise- and marginal-likelihood estimation for the mixed Rasch model with binary data. *Journal of Statistical Computation and Simulation* [IF=0.47], 82(3) :419–430, February 2012.
- [7] J. Golden, R. Michael Conroy, A. M. O'Dwyer, D. Golden, and J.-B. Hardouin. Illness-related stigma, mood and adjustment to illness in persons with hepatitis C. *Social Science & Medicine* [IF=2.74], 63(12) :3188–3198, December 2006. PMID : 17010490.
- [8] J. -B. Hardouin and M. Mesbah. The SAS Macro-Program %anaqol to estimate the parameters of Item Responses Theory models. *Communications in Statistics - Simulation and Computation* [IF=0.34], 36 :437–453, 2007.

- [9] J.-B. Hardouin. Rasch analysis : Estimation and tests with Raschtest. *Stata Journal* [IF=2.00], 7(1) :22–44, 2007.
- [10] J.-B. Hardouin, S. Amri, M.-L. Feddag, and V. Sébille. Towards power and sample size calculations for the comparison of two groups of patients with Item Response Theory models. *Statistics in Medicine* [IF=2.33], 31(11-12) :under press, May 2012. PMID : 22069169.
- [11] J.-B. Hardouin, A. Bonnaud-Antignac, and V. Sébille. Nonparametric Item Response Theory using Stata. *Stata Journal* [IF=2.00], 11(1) :30–51, 2011.
- [12] J.-B. Hardouin, R. Conroy, and V. Sébille. 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 Medical Research Methodology* [IF=2.15], 11 :105, 2011. PMID : 21756330.
- [13] J.-B. Hardouin and M. Mesbah. Clustering binary variables in subscales using an extended Rasch model and Akaike Information Criterion. *Communications in Statistics - Theory and Methods* [IF=0.35], 33 :1277–1294, December 2004.
- [14] M. Landreat-Guillou, C. Victorri-Vigneau, J.-B. Hardouin, M. Bronnec-Grall, M. Marais, J.-L. Venisse, and P. Jolliet. Can we say that seniors are addicted to benzodiazepines ? *Substance Use & Misuse* [IF=1.06], 45(12) :1988–1999, October 2010. PMID : 20465403.
- [15] C. Le Caignec, D. J Kwiatkowski, S. Küry, J.-B. Hardouin, J. Melki, and A. David. Three independent mutations in the TSC2 gene in a family with tuberous sclerosis. *European Journal of Human Genetics : EJHG* [IF=4.38], 17(9) :1165–1170, September 2009. PMID : 19259131.
- [16] E. Montassier, E. Batard, J. Segard, J.-B. Hardouin, A. Martinage, P. Le Conte, and G. Potel. Base excess is an accurate predictor of elevated lactate in ED septic patients. *The American Journal of Emergency Medicine* [IF=1.55], 30(1) :184–187, January 2012. PMID : 21159466.
- [17] P. Pottier, M. Fouassier, J.-B. Hardouin, C. Volteau, and B. Planchon. D-dimers, thrombin-antithrombin complexes, and risk factors for thromboembolism in hospitalized patient. *Clinical and Applied Thrombosis/Hemostasis : Official Journal of the International Academy of Clinical and Applied Thrombosis/Hemostasis* [IF=1.48], 15(6) :666–675, December 2009. PMID : 18796458.
- [18] P. Pottier, J.-B. Hardouin, T. Dejoie, A. Bonnaud, A.-G. Le Loupp, B. Planchon, and V. Leblanc. Stress responses in medical students in ambulatory and in-hospital patient consultations. *Medical Education* [IF=2.64], 45(7) :678–687, July 2011. PMID : 21649700.
- [19] P. Pottier, J.-B. Hardouin, B. D. Hodges, M.-A. Pistorius, J. Connault, C. Durant, R. Clairand, V. Sebille, J.-H. Barrier, and B. Planchon. Exploring how students think : a new method combining think-aloud and concept mapping protocols. *Medical Education* [IF=2.64], 44(9) :926–935, September 2010. PMID : 20716103.
- [20] P. Pottier, J.-B. Hardouin, S. Lejeune, P. Jolliet, B. Gillet, and B. Planchon. Immobilization and the risk of venous thromboembolism. A meta-analysis on epidemiological studies. *Thrombosis Research* [IF=2.37], 124(4) :468–476, September 2009. PMID : 19481781.
- [21] L. Roulet, F. Ballereau, J.-B. Hardouin, A. Chiffolleau, L. Moret, G. Potel, and N. Asseray. Assessment of adverse drug event recognition by emergency physicians

in a french teaching hospital. *Accepté pour publication dans Emergency Medicine Journal [IF=1.552]*, 2012.

- [22] V. Sébille, J.-B. Hardouin, T. Le Néel, G. Kubis, F. Boyer, F. Guillemin, and B. Falissard. 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 simulation study. *BMC Medical Research Methodology [IF=2.15]*, 10 :24, 2010. PMID : 20338031.
- [23] V. Sébille, J.-B. Hardouin, and M. Mesbah. Sequential analysis of latent variables using mixed-effect latent variable models : Impact of non-informative and informative missing data. *Statistics in Medicine [IF=2.33]*, 26(27) :4889–4904, November 2007. PMID : 17576119.
- [24] L. Wainstein, C. Victorri-Vigneau, V. Sébille, J.-B. Hardouin, F. Feuillet, J. Pivette, A. Chaslerie, and P. Jolliet. Pharmacoepidemiological characterization of psychotropic drugs consumption using a latent class analysis. *International Clinical Psychopharmacology [IF=2.76]*, 26(1) :54–62, January 2011. PMID : 20881845.

Articles dans des revues nationales et thèse de doctorat

- [25] J.-B. Hardouin. *Construction d'échelles d'items unidimensionnelles en qualité de vie*. PhD thesis, Paris Descartes - Paris V, Paris, November 2005.
- [26] C. Omnès and J.-B. Hardouin. Belote et rebelote, peut-on éviter le dix de der ? *Revue Francaise et Francophone de Psychiatrie et de Psychologie Médicale [IF inconnu]*, 76 :27–32, May 2004.
- [27] P. Pottier, B. Planchon, J.-B. Hardouin, V. Sébille, J.-M. Rogez, and J.-H. Barrier. Evaluation préliminaire de la validité de construit d'un dispositif cartographique (carte procédurale) pour l'étude du raisonnement clinique. *Pédagogie Médicale [IF inconnu]*, 2011.

Chapitres d'ouvrage

- [28] J.-B. Hardouin and M. Mesbah. Selection of items fitting a Rasch model. In *Mathematical methods for survival analysis, reliability and Quality of life*, pages 255–274. Huber C., Limnios N., Mesbah M., Nikulin M., Paris, hermes edition, 2008.
- [29] V. Sébille and J.-B. Hardouin. Sequential analysis of quality of life measurements using mixed Rasch models. In *Statistical models and methods for biomedical and technical systems*, pages 333–338. Vonta F., Nikulin M., Limnios N., Huber C., Boston, Birkäuser edition, 2008.

Articles en révision

- [30] J.-F. Hamel, J.-B. Hardouin, T. Le Néel, G. Kubis, Serge Fanello, and V. Sébille. Study of different methods for comparing groups by analysis of subjective health measurements. *En révision dans PLoS ONE [IF=4.41]*, 2012.

Articles soumis ou en cours d'écriture

- [31] E. Batard, F. Ollivier, D. Boutoille, J.-B. Hardouin, E. Montassier, J. Caillon, and F. Ballereau. Relationship between hospital antibiotic use and quinolone resistance in escherichia coli. *Soumis*, 2012.
- [32] M. Blanchin, J.-B. Hardouin, A. Guilleux, F. Guillemin, B. Falissard, and V. Sébille. Validation of a method to determine the power of a test comparing groups of patients with a Rasch model. *En cours d'écriture*, 2012.
- [33] M. Blanchin, J.-B. Hardouin, T. Le Néel, G. Kubis, and V. Sébille. CTT and IRT-based approaches for group comparison of longitudinal patient reported outcomes : Impact of dropout mechanisms. *Soumis*, 2011.
- [34] G. Bouju, J.-B. Hardouin, C. Boutin, P. Gorwood, J.-D. Le Bourvellec, F. Feuillet, J.-L. Vénisse, and M. Bronnec. A shorter and multidimensional version of the gambling beliefs and attitudes survey (GABS-23). *Soumis*, 2011.
- [35] P. Clermidi, G. Podevin, C. Crétole, Sarnacki S., and J.-B. Hardouin. The challenge of measuring quality of life in children with Hirschsprung's diseases or anorectal malformation. *En cours d'écriture*, 2011.
- [36] E. Duclay, J.-B. Hardouin, V. Sébille, and L. Moret. Impact of staff absenteeism on patient satisfaction : an exploratory study in a french hospital. *En cours d'écriture*, 2012.
- [37] E. Dumas, J.-B. Hardouin, M. Blanchin, T. Le Néel, G. Kubis, and V. Sébille. Analysis of longitudinal Patient-Reported outcomes with intermittent missing data (informative and non informative) : comparison of score or Rasch-based methods. *En cours d'écriture*, 2011.
- [38] F. Feuillet, L. Bellanger, J.-B. Hardouin, C. Victorri-Vigneau, and V. Sébille. On comparison of clustering methods for pharmacoepidemiological data. *Soumis*, 2012.
- [39] J.-F. Hamel, V. Sébille, T. Le Néel, G. Kubis, and J.-B. Hardouin. Rasch models to detect a group effect for Patients Reported Outcomes with possibly informative missing data. *Soumis*, 2011.
- [40] J.-B. Hardouin, E. Audureau, A. Leplège, and J. Coste. Spatio-temporal Rasch analysis of quality of life outcomes in the french general population. measurement invariance and group comparisons. *Soumis*, 2011.
- [41] J.-B. Hardouin, M. Blanchin, M. Feddag, B. Falissard, and V. Sébille. Adaptation of a method to determine the power of a test comparing groups of patients to a polytomous Rasch model. *En cours d'écriture*, 2012.
- [42] J.-B. Hardouin, F. Feuillet, C. Crétole, Sarnacki S., and Podevin G. Validation of the french versions of the Hirschsprung's disease and Anorectal malformations Quality of Life questionnaire for adults. *En cours d'écriture*, 2011.
- [43] J.-B. Hardouin, B. Planchon, and P. Pottier. Validation of a questionnaire measuring the clinical reasoning. *En cours d'écriture*, 2012.
- [44] E. Montassier, E. Batard, J.-B. Hardouin, J. Caillon, S. Le Fresne, T. Carton, N. Caroff, T. Gastinne, P. Moreau, G. Potel, F. Le Vacon, and M.-F. De La CochetiÁÍre. Chemotherapy-induced changes of the intestinal microbiota analyzed by high throughput pyrosequencing. *Soumis*, 2012.

- [45] P. Pottier, T. Dejoie, J.-B. Hardouin, A.-G. Le Loupps, B. Planchon, A. Bonnaud-Antignac, and V. Leblanc. Stressful conditions in simulated ambulatory consultations generate a high stress without modifying decision-making. *Soumis*, 2011.
- [46] L. Roulet, F. Ballereau, J.-B. Hardouin, A. Chiffolleau, G. Potel, and N. Asseray. Adverse drug event non-recognition in emergency departments : Factors related to patients and drugs. *Soumis*, 2012.
- [47] V. Sébille, M. Blanchin, F. Guillemin, B. Falissard, and J.-B. Hardouin. Required sample size in order to compare groups of patients with a Rasch model. *En cours d'écriture*, 2012.

Communications internationales

- [48] M. Blanchin, J.-B. Hardouin, T. Le Néel, G. Kubis, and V. Sébille. Comparison of three methods for the analysis of longitudinal patient reported outcomes. In *16th International Meeting of the Psychometric Society (IMPS)*, Cambridge, 2009.
- [49] M. Blanchin, J.-B. Hardouin, T. Le Néel, G. Kubis, and V. Sébille. Comparison of three methods for the analysis of longitudinal patient reported outcomes. In *Probabilistic models for measurement in education, psychology, social science and health*, Copenhagen, 2010.
- [50] J.-F. Hamel, J.-B. Hardouin, T. Le Néel, G. Kubis, and V. Sébille. Rasch models to detect a group effect for patients reported outcomes with possibly informative missing data. In *International Workshop on "Patient Reported Outcomes and Quality of Life" - Annales de l'ISUP*, volume 55, pages 49–52, Paris, 2011.
- [51] J.-B. Hardouin. Construction of QoL subscales using a multidimensional item response model. In *International Workshop "Environment and Health Related Quality of Life : Statistical Perspectives"*, page 21, Vannes, 2002.
- [52] J.-B. Hardouin. A process of items selection in subscales using a parametric multidimensional IRT model. In *23rd Biennial Conference of the Society for Multivariate Analysis in the Behavioural Sciences (SMABS)*, pages 68–69, Tilburg, The Netherlands, 2002.
- [53] J.-B. Hardouin. Classification of items into quality of life sub-scales. In *International Conference on : Statistics in Health Sciences*, pages 151–154, Nantes, 2004.
- [54] J.-B. Hardouin. Clustering of quality of life items around latent variables. In *International Conference on Statistical Latent Variables Models in the Health Sciences*, pages 57–60, Perugia, Italy, 2006.
- [55] J.-B. Hardouin. Item response theory with the SAS software : choice of a model and selection of quality of life items. In *European Seminar on Mathematical Methods in Survival Analysis, Reliability and Quality of Life*, Paris, 2006.
- [56] J.-B. Hardouin, L. Azema, B. Planchon, N. Asseray, M. Miegerville, J.-H. Barrier, and P. Pottier. Validation d'un questionnaire d'évaluation des stratégies de résolution de problème. In *Premier Congrès International Francophone de Pédagogie des Sciences de la Santé et et Sport*, Grenoble, 2009.
- [57] J.-B. Hardouin, M. Blanchin, M. -L. Feddag, T. Le Néel, G. Kubis, A. Bonnaud-Antignac, P. Tessier, F. Guillemin, B. Falissard, and V. Sébille. Sample size requirement for the comparison of groups of patients with dichotomous and polytomous Rasch models : validation and robustness of an adapted formulae. In

- Statistical Models and Methods for Reliability and Survival Analysis and their Validation (S2MRSa)*, Bordeaux, 2012.
- [58] J.-B. Hardouin, B. Falissard, T. Le Néel, G. Kubis, F. Guillemin, and V. Sébille. Comparison of classical test and item response theories in health. In *31st annual conference of the international society of clinical biostatistics (ISCB)*, Montpellier, 2010.
- [59] J.-B. Hardouin, B. Falissard, T. Le Néel, G. Kubis, and V. Sébille. Power with CTT and IRT based approaches for the comparison of two groups of patients. In *16th International Meeting of the Psychometric Society (IMPS)*, Cambridge, 2009.
- [60] J.-B. Hardouin, A. Leplège, E. Audureau, and J. Coste. Spatio-temporal Rasch analysis of quality of life outcomes in the french general population. In *Probabilistic models for measurement in education, psychology, social science and health*, Copenhagen, 2010.
- [61] J.-B. Hardouin and M. Mesbah. A process of items selection in subscales using a parametric multidimensional IRT model. In *2nd Euro-Japanese Workshop on Stochastic Modelling for Finance, Insurance, Production and Reliability*, pages 128–137, Chamonix, 2002.
- [62] J.-B. Hardouin and M. Mesbah. Classification of ordinal variables measuring the same latent trait using IRT models and GEE. In *Workshop on Probability, Statistics and Modelling in Public Health. In honour of Professor Marvin Zelen (Harvard University)*, Bordeaux, 2003.
- [63] J.-B. Hardouin and V. Sébille. Impact of imputation methods for missing data for the analysis of subjective health outcome measures. In *International Workshop on Advances in Rasch Modelling*, Vannes, 2007.
- [64] P. Pottier, J.-B. Hardouin, T. Dejoie, B. Planchon, J.-M. Rogez, and V. Leblanc. Physiological and psychological responses to stress in year 6 medical students faced to ambulatory symptomatic patients in internal medicine. In *AMEE 2010 (Association for medical european education)*, Glasgow, 2010.
- [65] V. Sébille, S. Amri, and J.-B. Hardouin. Towards sample size determination for IRT analysis for the comparison of two groups of patients. In *Probabilistic models for measurement in education, psychology, social science and health*, Copenhagen, 2010.
- [66] V. Sébille, M. Blanchin, M.-L. Feddag, S. Amri, T. Le Néel, G. Kubis, F. Guillemin, B. Falissard, and J.-B. Hardouin. Methodological issues regarding the planning and the analyses of pro related studies. In *6th Congress of the Italian Society of Medical Statistics and Clinical Epidemiology (SISMEC)*, Ancona, Italia, 2011.
- [67] V. Sébille, B. Falissard, S. Amri, M.-L. Feddag, T. Le Néel, G. Kubis, F. Guillemin, and J.-B. Hardouin. Sample size determination for IRT. why should it differ from CTT? In *International Workshop on "Patient Reported Outcomes and Quality of Life" - Annales de l'ISUP*, volume 55, pages 73–74, Paris, 2011.
- [68] V. Sébille and J.-B. Hardouin. Item response theory models and early stopping of clinical trials with quality of life measurements. In *International Conference on Statistical Latent Variables Models in the Health Sciences*, pages 96–96, Perugia, Italy, 2006.

Posters dans des conférences internationales

- [69] M. Blanchin, J.-B. Hardouin, T. Le Néel, G. Kubis, and V. Sébille. Analysis of longitudinal patient reported outcomes with missing data. In *31st annual conference of the international society of clinical biostatistics (ISCB)*, Montpellier, 2010.
- [70] E. Dumas, J.-B. Hardouin, M. Blanchin, T. Le Néel, G. Kubis, and V. Sébille. Analysis of longitudinal Patient-Reported outcomes with intermittent missing data (informative and non informative) : comparison of score or Rasch-based methods. In *Annales de l'ISUP - International Workshop on "Patient Reported Outcomes and Quality of Life"*, volume 55, pages 43–46, Paris, 2011.
- [71] J.-F. Hamel, J.-B. Hardouin, T. Le Néel, G. Kubis, and V. Sébille. Rasch models to detect a group effect for patients reported outcomes. In *31st annual conference of the international society of clinical biostatistics (ISCB)*, Montpellier, 2010.
- [72] J.-B. Hardouin, L. Azema, B. Planchon, N.. Asseray, M. Miegerville, J.-H. Barrier, and P. Pottier. Validation d'un questionnaire d'évaluation des stratégies de résolution de problèmes chez des étudiants en médecine pré-gradués. phase III. In *5ème Forum International francophone de pédagogie des sciences de la santé*, 2010.
- [73] J.-B. Hardouin, B. Falissard, T. Le Néel, G. Kubis, and V. Sébille. Power with CTT and IRT based approaches for the comparison of two groups of patients. In *31st annual conference of the international society of clinical biostatistics (ISCB)*, Montpellier, 2010.
- [74] E. Montassier, E. Batard, J.-B. Hardouin, M. Miegerville, J. Caillon, N. Caroff, S. Le Fresne, T. Carton, T. Gastinne, P. Moreau, G. Potel, F. Le Vacon, and M.-F. De la Cochetière. A specific intestinal microbiota profile predisposes to severe chemotherapy-induced diarrhoea. In *22nd European Congress of Clinical Microbiology and Infectious Diseases (ECCMID 2012)*, Londres, 2012.
- [75] P. Pottier, J.-B. Hardouin, T. Dejoie, M. Flori, B. Planchon, J.-M. Rogez, and V. Leblanc. Réponses physiologiques et psychologiques au stress chez des étudiants en 6ème année de médecine confrontés à des problèmes diagnostiques en milieu ambulatoire. In *5ème Forum International francophone de pédagogie des sciences de la santé*, 2010.

Communications nationales

- [76] E. Audureau, S. Rican, J.-B. Hardouin, and J. Coste. Qualité de vie liée à la santé : dégradation globale et accroissement des disparités en France entre 1995 et 2003. In *5ème Conférence Francophone d'Epidémiologie Clinique*, page S5o04, Marseille, 2011.
- [77] M. Blanchin, J.-B. Hardouin, T. Le Néel, G. Kubis, and J.-B. Sébille. Comparaison de trois méthodes d'analyse des variables latentes longitudinales. In *41èmes Journées de Statistique*, Bordeaux, 2009.
- [78] F. Boyer, A. Calmus, C. Reveillère, J.-L. Novella, I Morrone, M Espagnacq, A Wolak, C. Barbe, M. Dramé, J.-B. Hardouin, E Le Rohellec, I. Richard, V. Tiffreau, F Blanchard, and D. Jolly. Les mesures subjectives rapportées par les patients ("la qualité de vie liée à la sante" ou "Patient reported outcome") : concepts, applications. In *XXIIIe entretiens de la fondation Garches - Handicap et qualité*

- de vie*, Editions GMSanté, pages 75–84, Neuilly sur Seine, 2010. J.-F. Ravaud, I. Ville et S. Poiraudau.
- [79] F. Boyer, C. Reveillère, A. Calmus, V. Tiffreau, I. Richard, G Bassez, Yann Péréon, J.-B. Hardouin, Sophie Minh-Muzeaux, J.-L. Novella, C. Barbe, A Wolak, I Morrone, M. Dramé, and D. Jolly. Développement d’un outil spécifique pour les maladies neuromusculaires (MNM) : le QoL-NMD. génération qualitative et quantitative de la banque d’items. In *XXIIIe entretiens de la fondation Garches - Handicap et qualité de vie*, Editions GMSanté, pages 59–66, Neuilly sur Seine, 2010. J.-F. Ravaud, I. Ville et S. Poiraudau.
- [80] F. Boyer, V. Tiffreau, I. Richard, A. Calmus, L. Percebois-Macadre, J.-B. Hardouin, C. Reveillère, M. Dramé, I. Morrone, C. Barbe, A. Wolak, J.-L. Novella, and D. Jolly. Développement d’un outil spécifique pour MNM : le QoL-NMD. génération qualitative et quantitative de la banque d’items. In *26e congrès de médecine physique et de réadaptation*, 2011.
- [81] J.-B. Hardouin. Les patients vus au centre hospitalier de Dreux en 2000 pour tentative de suicide : typologie des patients & prédiction du risque de récurrence. In *Conférence pour les 10 ans de l’Observatoire Régional de la Santé (ORS) du Centre*, Orléans, 2007.
- [82] J.-B. Hardouin, I. Le Houérou, and E. Mounier. Enquête descriptive sur les patients accueillis dans les structures de soin en alcoologie d’Eure-et-Loir. In *6e forum européen de l’alcoologie plurielle*, Hénin-Beaumont, 2003.
- [83] J.-B. Hardouin, I. Le Houérou, and E. Mounier. Evaluation du devenir des patients pris en charge par les structures de soin d’Eure-et-Loir. In *Réunion de la Société Française d’Alcoologie*, Paris, 2004.
- [84] J.-B. Hardouin and J.-B. Sébille. Impact des méthodes d’imputation de données manquantes pour l’analyse de mesures subjectives en santé. In *39e journées de la Statistique*, Angers, 2007.
- [85] J.-B. Hardouin and J.-B. Sébille. Comparaison de méthodes d’imputation de données manquantes pour l’analyse de mesures subjectives en santé. In *Biostatistique et mesures subjectives en santé*, Nantes, 2008.
- [86] C. Omnès and J.-B. Hardouin. Belote et rebelote, est-il possible de prédire avant le dix de der ? In *XXXVèmes Journées du Groupement d’Etude sur la Prévention du Suicide (GEPS)*, Nancy, 2003.
- [87] C. Omnès and J.-B. Hardouin. Rencontre du 5e type. In *3e congrès international francophone de prévention du suicide*, Poitiers, 2004.
- [88] V. Sébille, J.-B. Hardouin, and A. Bonnaud-Antignac. Biostatistique & qualité de vie en cancérologie : étude longitudinale de la qualité de vie et des stratégies de coping des patientes atteintes de cancer du sein et de leur entourage. In *Qualité de vie et cancer : pourquoi ? Comment ?*, Nantes, 2007. Cancéropole Grand ouest.

Posters dans des conférences nationales

- [89] M. Blanchin, J.-B. Hardouin, T. Le Néel, G. Kubis, A. Bonnaud-Antignac, and V. Sébille. Comparison of three methods for the analysis of longitudinal patient reported outcomes. In *Biostatistique et mesures subjectives en santé*, Nantes, 2008.

- [90] M. Blanchin, J.-B. Hardouin, T. Le Néel, G. Kubis, and V. Sébille. Comparison of three methods for the analysis of longitudinal patient reported outcomes. In *Epidémiologie et Biométrie 2008*, 2009.
- [91] A. Bonnaud-Antignac, J.-B. Hardouin, M. Quere, F. Feuillet, A. Louedec, A. Ouadi, F. Nazih, S. Biacabe, A.-H. Boudoukha, E. Boujut, L. Dréno, A. Jean, Merle C., V. Sebille, R. Senand, P. Tessier, A. Florin, A. Sarrazin, C. Burban, O. Laboux, A. Pineau, and J.-M. Rogez. Etude du vécu des étudiants en première année commune aux études de santé (PAES) de l'université de Nantes - Approche quantitative. In *Les SHS au chevet de la médecine*, Nantes, 2011.
- [92] M. Bronnec, G. Bouju, J.-B. Hardouin, P. Gorwood, Morgane Guillou-Landreat, and Jean Luc Venisse. Le GABS et le GRCS : 2 questionnaires utiles pour évaluer la sévérité des distorsions cognitives du joueur pathologique. In *7e congrès de l'Encéphale*, 2009.
- [93] F. Feuillet, J.-B. Hardouin, Laura Wainstein, Caroline Victorri-Vigneau, Pascale Jolliet, and V. Sébille. Caractérisation pharmaco-épidémiologique de patients consommateurs de médicaments psychotropes : Apport des modèles à classes latentes. In *Recherche en Pharmaco-épidémiologie*, Nantes, 2009.
- [94] J.-B. Hardouin, T. Le Néel, G. Kubis, and V. Sébille. Comparaison de mesures issues de Patients Reported Outcomes entre deux groupes de patients : identification de facteurs influant sur la puissance. In *Biostatistique et mesures subjectives en santé*, Nantes, 2008.
- [95] J.-D. Le Bourvellec, M. Bronnec, G. Bouju, and J.-B. Hardouin. Mesure des distorsions cognitives des joueurs pathologiques : validation en langue française des questionnaires GABS et GRCS. In *Recherche en Pharmaco-épidémiologie*, Nantes, 2009.
- [96] J. Léger, J.-B. Hardouin, A. Bonnaud-Antignac, and V. Sébille. Analyse longitudinale de données de qualité de vie de patientes atteintes de cancer du sein : influence du coping et des caractéristiques des patientes. In *Biostatistique et mesures subjectives en santé*, Nantes, 2008.
- [97] E. Montassier, E. Batard, J. Segard, A. Martinage, J.-B. Hardouin, P. Le Conte, and G. Potel. L'excès de base : un marqueur sûr pour prédire l'élévation du taux de lactate chez les patients septiques admis aux urgences. In *Urgences 2011*, 2011.
- [98] P. Pottier, N. Asseray, A. Bonnaud-Antignac, R. Clément, G. Dabouis, L. Goronflot, C. Greffier, J.-B. Hardouin, J. Le Mauff, M. Miegerville, C. Rat, V. Sébille, R. Senand, J.-M. Rogez, B. Planchon, and J.-H. Barrier. Création et développement d'une structure de recherche en éducation médicale à la faculté de médecine de Nantes : l'ERNEM. In *Journées Francophones de Pédagogie en Sciences de la Santé*, Lille, 2008.
- [99] P. Pottier, J.-B. Hardouin, B. Planchon, V. Sébille, and J.-H. Barrier. Validation d'un nouvel outil d'évaluation individuelle des stratégies de résolution de problème basé sur la rédaction de cartes conceptuelles chez les étudiants en 5ème année de médecine. In *Biostatistique et mesures subjectives en santé*, Nantes, 2008.

Macro-programmes SAS

- [100] J.-B. Hardouin. %Backrasch : A SAS macro-program to select items fitting a Rasch model with backward procedure, June 2004.

- [101] J.-B. Hardouin. %Detect : a SAS macro-program to compute the DETECT, ISS and R indexes, May 2004.
- [102] J.-B. Hardouin. %Gammasym : a SAS macro-program to estimate the Symetrical gamma function, February 2004.
- [103] J.-B. Hardouin. %Gausshermite : a SAS macro-program to approximate an integral with Gauss-Hermite quadrature, May 2005.
- [104] J.-B. Hardouin. %LoevH : A SAS macro-program to realize a Mokken analysis, February 2007.
- [105] J.-B. Hardouin. %AnaQol : A SAS macro-program to estimate the parameters and fit of IRT models, October 2009.
- [106] J.-B. Hardouin. %MSP : A SAS macro-program to realize a Mokken Scale Procedure, July 2011.

Modules Stata

- [107] J.-B. Hardouin. Detect : A Stata module to compute the DETECT, ISS and R indexes, May 2004.
- [108] J.-B. Hardouin. Backrasch : A Stata module to select items in a Rasch model with backward procedure, May 2005.
- [109] J.-B. Hardouin. Gausshermite : A Stata module to approximate an integral with Gauss-Hermite quadrature, May 2005.
- [110] J.-B. Hardouin. Hcavar : A Stata module to perform a Hierarchical Cluster Analysis on variables, December 2005.
- [111] J.-B. Hardouin. MMSRM : A Stata module to fit a Multidimensional Marginally Sufficient Rach Model, July 2005.
- [112] J.-B. Hardouin. CLV : A Stata module to perform a Classification around Latent Variables, December 2006.
- [113] J.-B. Hardouin. Raschfit : A Stata module to perform the Raschfit and Raschfit-Fast algorithms, January 2006.
- [114] J.-B. Hardouin. Simirt : A Stata module to simulate data with IRT models, October 2006.
- [115] J.-B. Hardouin. Gammasym : a Stata module to estimate the symetrical Gamma function, March 2007.
- [116] J.-B. Hardouin. Trace : A Stata module to display trace-lines of items, May 2007.
- [117] J.-B. Hardouin. Imputeitems : A Stata module to impute items with classical methods, November 2008.
- [118] J.-B. Hardouin. Imputemok : A Stata module to impute items with the Mokken algorithm, November 2008.
- [119] J.-B. Hardouin. Imputerasch : A Stata module to impute items with the Rasch model, June 2008.
- [120] J.-B. Hardouin. Loevh : A Stata module to perform a Mokken analysis, November 2008.
- [121] J.-B. Hardouin. Raschtest : A Stata module to fit a Rasch model, June 2008.

Rapports

- [122] C. Cohonner, J.-B. Hardouin, and C. Leclerc. Evaluation du réseau DIABOLO - volet économique. Rapport d'étude, ORS du Centre, Orléans, 2006.
- [123] C. Cohonner, J.-B. Hardouin, and C. Leclerc. Evaluation du réseau DIABOLO - volet patients. Rapport d'étude, ORS du Centre, Orléans, 2006.
- [124] C. Cohonner, J.-B. Hardouin, A. Pouliquen, and C. Leclerc. Exploitation de l'enquête "urgences en médecine libérale". Rapport d'étude, ORS du Centre, Orléans, 2006.
- [125] N. De Courcel, J.-B. Hardouin, and C. Leclerc. Projet d'établissement 2007-2011 : Les caractéristiques sanitaires de la zone d'emploi de Châteaudun. Rapport d'étude, ORS du Centre, Orléans, 2006.
- [126] M. Duval, J.-B. Hardouin, and M. Boin. Epidémiologie des cancers en région Centre (1980-2000). Rapport d'étude, ORS du Centre, ARH du Centre, Orléans, 2004.
- [127] G. Falletta, J.-B. Hardouin, and F. Kasiborski. Réalisation de groupes homogènes de malades (GHM). Rapport de stage de licence professionnelle en génie informatique et statistique (GIS), Département d'Informatique Médicale (DIM) du Centre Hospitalier de Vannes, 1997.
- [128] J.-B. Hardouin. La pertinence des alarmes en service de réanimation adulte. Rapport de stage de DUT statistique et traitement informatique des données (STID), Laboratoire de Biomathématiques - Université de Lille 2, 1996.
- [129] J.-B. Hardouin. Les modèles linéaires généralisés : une application pour prédire la présence et l'abondance de quelques mauvaises herbes en Côte d'Or. Rapport de stage de maîtrise professionnelle en génie informatique et statistique (GIS), Laboratoire de Malherbologie - INRA Dijon, 1998.
- [130] J.-B. Hardouin. Capture-recapture : l'apport du Robust Design en Épidémiologie. Rapport de stage de DEA santé publique - biostatistique, Laboratoire de Statistique Appliquée de Bretagne Sud (SABRES) - Université de Bretagne Sud, 1999.
- [131] J.-B. Hardouin and M. Boin. Etat des lieux dans la région Centre en vue de la mise en place de la campagne de dépistage organisé du cancer du sein. Rapport d'étude, ORS du Centre, Orléans, 2001.
- [132] J.-B. Hardouin and M. Boin. La prise en charge des personnes suicidaires dans la région Centre. Rapport d'étude, ORS, Orléans, 2002.
- [133] J.-B. Hardouin and M. Boin. Analyse des certificats de santé du 8ème jour des enfants nés en région Centre en 2000. Rapport d'étude, ORS du Centre, Orléans, 2003.
- [134] J.-B. Hardouin and M. Boin. Analyse des certificats de santé du 8ème jour des enfants nés en région Centre en 2002. Rapport d'étude, ORS du Centre, Orléans, 2004.
- [135] J.-B. Hardouin and M. Boin. Démographie médicale et inégalités cantonales en région Centre. Rapport d'étude, ORS du Centre, Orléans, 2004.
- [136] J.-B. Hardouin, M. Boin, E. Mounier, and I. Le Houérou. Enquête descriptive sur les patients accueillis dans les structures de soin en alcoologie d'Eure-et-Loir : l'entrée et la sortie des cures. Rapport d'étude, ORS du Centre, Orléans, 2004.

- [137] J.-B. Hardouin, M. Boin, E. Mounier, and I. Le Houérou. Enquête descriptive sur les patients accueillis dans les structures de soin en alcoologie d'Eure-et-Loir : un an après la cure. Rapport d'étude, ORS du Centre, Orléans, 2004.
- [138] J.-B. Hardouin, M. Boin, and C. Omnès. Profil des personnes ayant réalisé une tentative de suicide au centre hospitalier général de Dreux. Rapport d'étude, ORS du Centre, Orléans, 2004.
- [139] J.-B. Hardouin and C. Leclerc. Analyse des certificats de santé du 8ème jour des enfants nés en région Centre en 2003. Rapport d'étude, ORS du Centre, Orléans, 2005.
- [140] J.-B. Hardouin and C. Leclerc. Le stress chez les personnes âgées de 60 ans et plus à Orléans : causes et impact quotidien. Rapport d'étude, ORS du Centre, Orléans, 2005.
- [141] J.-B. Hardouin, M. Pichard, and C. Leclerc. Evaluation du réseau santé ou tabac d'Eure-et-Loir. Rapport d'étude, ORS du Centre, Orléans, 2005.
- [142] J.-B. Hardouin, A. Pouliquen, and C. Leclerc. Analyse des certificats de santé du 8ème jour des enfants nés en région Centre en 2004. Rapport d'étude, ORS du Centre, Orléans, 2006.
- [143] J.-B. Hardouin and A. Viallefont. A new design for estimating prevalence of diseases by Capture-Recapture. Technical Report hal-00020083_v1, Laboratoire de Statistique Appliquée de Bretagne Sud (SABRES) - Université de Bretagne Sud, 2001.

ENCADREMENT DOCTORAL ET SCIENTIFIQUE

Encadrement doctoral

- **Myriam Blanchin** (coencadrement à 50%) [janvier 2008 - juin 2011] *Développement de modèles de réponse aux items (modèles IRT) pour l'analyse de mesures subjectives en santé évaluées dans le cadre d'études longitudinales*
 - Revues internationales : 2 articles publiés [1, 2], 1 article soumis [33]
 - Communications orales : 2 internationales [48, 49], 1 nationale [77]
 - Posters : 2 internationaux [69, 70], 2 nationaux [89, 90]
- **Jean-François Hamel** (M2 Modélisation en Pharmacologie Clinique et Epidémiologie (Nantes) et doctorat : coencadrement à 60%) [doctorat débuté en octobre 2010] *Etude de différentes approches permettant la comparaison de groupes d'individus par l'analyse de mesures subjectives en santé en présence de données manquantes potentiellement informatives*
 - Revues internationales : 1 article en revision [30] et 1 article soumis [39]
 - Communications orales : 1 internationale [50]
 - Posters : 1 international [71]
- **Elodie De Bock-Dumas** (coencadrement à 60%) [doctorat débuté en janvier 2011] *Identification de stratégies d'analyse de variables latentes longitudinales en présence de données manquantes potentiellement informatives*
 - Revues internationales : 1 article en cours d'écriture [37]
 - Posters : 1 international [70]

Accompagnement scientifique de doctorant (sans statut d'encadrant)

- **Fanny Feuillet** [doctorat débuté en octobre 2011] *Apport des analyses en variables latentes en pharmaco-épidémiologie*
 - un article soumis [38]
- **Gaëlle Bouju** [doctorat débuté en octobre 2011] *Les jeux de hasard et d'argent : vers une approche spécifique des différentes formes de jeux*
 - un article soumis [34]
- **Emmanuelle Anthoine** [doctorat débuté en octobre 2011] *Détermination du Nombre de Sujets Nécessaires (NSN) dans le cadre de la validation classique d'un score type Patient Reported Outcomes (PRO).*

Encadrement de Master 2 avec production scientifique

- **Estelle Duclay** (coencadrement à 40%) [Master 2 Modélisation en Pharmacologie Clinique et Epidémiologie (Nantes) - 2011] *Influence de l'absentéisme des personnels soignants sur la satisfaction du patient hospitalisé*
 - 1 article en cours d'écriture [36]
- **Pauline Clermidi** (coencadrement à 80%) [Master 2 Modélisation en Pharmacologie Clinique et Epidémiologie (Nantes) 2010] *Validation de la traduction française du questionnaire de qualité de vie spécifique HAQL auprès de patients de 6 à 11 ans atteints de malformations ano-rectales*
 - 1 article en cours d'écriture [35]
- **Fanny Feuillet** (coencadrement à 50%) [Master de Biostatistique (Bordeaux) 2009] et **Laura Wainstein** (coencadrement à 20%) [Master 2 Méthodologie et Statistique en Recherche Biomédicale (Paris 11) 2009] *Caractérisation pharmaco-épidémiologique de patients consommateurs de médicaments psychotropes : Apport des modèles à classes latentes*
 - 1 article publié dans une revue internationale [24], 1 poster dans un colloque national [93]
- **Julie Léger** (Master de Biostatistique (Bordeaux) coencadrement à 50%) [2008] *Analyse longitudinale de données de qualité de vie de patientes atteintes du cancer du sein : influence du coping et des caractéristiques des patientes*
 - 1 article dans une revue internationale [3], 1 poster dans un colloque national [96]

Autres encadrements d'étudiants avec production scientifique

- **Jean-Damien Le Bourvellec** (encadrement à 100%) [Licence Pro Statistique et Informatique pour la Santé (Bretagne Sud)- 2009] *Mesure des distorsions cognitives des joueurs pathologiques : validation en langue française des questionnaires GABS et GRCS*
 - 1 article dans une revue internationale [4], 1 article soumis dans une revue internationale [34], 1 poster dans un colloque national [95]
- **Sarah Amri** (coencadrement à 50%) [ENS Cachan - Antenne de Bretagne (Bruz) - 2009] *Estimation d'un effet groupe dans le cadre du modèle de Rasch à effet aléatoire : approche numérique de l'erreur standard de l'estimation*

- 1 article publié dans une revue internationale [10]
- 3 communications internationales [65, 66, 67]

Autres encadrements prévus ou n'ayant pas encore donné lieu à une valorisation scientifique

- **Mathilde Quere** (coencadrement à 100%) [DUT STID (Bretagne Sud) - 2011] *Analyse d'un questionnaire de mesure du stress et de la qualité de vie des étudiants de première année commune aux études de santé*
- **Alexandra Rouquette** (encadrement à 100%) [thèse d'exercice de médecine (Angers)- 2012] *Détermination d'une différence cliniquement pertinente sur le questionnaire GABS23 (Gambling Attitudes and Beliefs Scale - 23 items) auprès d'un échantillon de 609 joueurs*
- **Emilie Peron** (coencadrement à 40%) [Master 2 Modélisation en Pharmacologie Clinique et Epidémiologie (Nantes) - 2012] *Facteurs déterminants de la prescription de 6 IRSS pour des patients naïfs de tout traitement antidépresseur*
- **Gaëlle Romain** (coencadrement à 50%) [Licence Pro Statistique et Informatique en Santé (Bretagne Sud)- 2012] *Etude d'une cohorte de 623 joueurs de jeux de hasard et d'argent : analyses des différentes formes de jeux*

RAYONNEMENT SCIENTIFIQUE

Conférences invitées

Liste des conférences invitées :

- Conférences internationales : [52, 53, 54, 55, 57, 63, 65, 66, 67]
- Conférences nationales : [82, 83]

Co-responsabilité de contrats publics ou privés

- *Méthodes pour la planification des études portant sur des mesures subjectives en santé (PLANIPRO)* [2011-2013]
- **Coordinateur associé** (Coordinateur principal : Dr Véronique Sébille)
- Financement ANR 2009 dans le cadre du Programme de Recherche en Santé Publique (PRSP)

Participation à des contrats comme méthodologiste/statisticien

Agence Nationale de la Recherche (ANR)

- ANR - Programme SAMENTA 2012 : *Etude du profil des malades de Parkinson développant une addiction de type comportementale au cours d'un traitement dopaminergique : aspects psychopathologiques, neurologiques et pharmacocinétiques (PARKADD)* [Investigateur principal : Dr Marie-Grall-Bronnec (CHU Nantes)]

Agence Française de Sécurité Sanitaire des Produits de Santé (afssaps)

- Programme annuel 2012 : *Création d'un outil de caractérisation des pharmacodépendances médicamenteuses (PROMESS)* [Investigateur principal : Dr Caroline Victorri-Vigneau (CHU Nantes)]

Programme de Soutien aux Techniques Innovantes Coûteuses (STIC)

- Financement 2011 : *Evaluation médico-économique de la fermeture des plans sous-cutanés par l'utilisation d'une agrafeuse à agrafes résorbables (INSORB) (S2CARA)* [Investigateur principal : Pr Olivier Malard (CHU Nantes)]
- Financement 2010 : *Evaluation médico-économique des nouveaux traitements chirurgicaux de la maladie hémorroïdaire : Essai randomisé multicentrique comparant la ligature élective des artères hémorroïdales sous contrôle doppler avec mucopexie par voie trans-anales à l'hémorroïdopexie circulaire par agrafage selon la technique de Longo (LIGALONGO)* [Investigateur principal : Pr Paul-Antoine Lehur (CHU Nantes)]
- Financement 2010 : *Evaluation médico-économique de la surveillance par oxymétrie cérébrale (INVOS) en peropératoire lors d'endartériectomies carotidiennes (EMOCAR)* [Investigateur principal : Dr Yann Le Teurnier (CHU Nantes)]

Programmes Hospitaliers de Recherche Clinique (PHRC)

- En cours d'évaluation : *Modérateurs de jeux en ligne : quelle efficacité ? Etude expérimentale contrôlée et randomisée (MOD&JEU)* [Investigateur principal : Pr Jean-Luc Vénisse (CHU Nantes)]
- Financement 2012 : *Etude multicentrique, randomisée, comparant 2 stratégies pour diminuer la consommation d'antibiotiques dans les pneumopathies aiguës communautaires : une stratégie basée sur la réévaluation clinique et une stratégie guidée par la PCT* [Investigateur principal : Dr Emmanuel Montassier (CHU Nantes)]
- Financement 2011 : *Etude clinique monocentrique, prospective, comparant deux groupes de joueurs problématiques (" jeux en ligne " versus " jeux hors ligne ") débutant des soins dans le service d'Addictologie du CHU de Nantes (EVALJEU-en ligne)* [Investigateur principal : Dr Marie-Grall-Bronnec (CHU Nantes)]
- Financement 2010 : *Etude multicentrique nationale de la qualité de vie des patients atteints de malformation anorectale (MAR-QoL)* [Investigateur principal : Pr Guillaume Podevin (CHU Angers)]
- Financement 2009 : *Suivi d'une cohorte française de joueurs (jeux de hasard et d'argent). Evolution de la pratique de jeu et recours aux soins (EVALJEU)* [Investigateur principal : Dr Marie Bronnec (CHU Nantes)]
- Financement 2007 : *Conception et validation d'un questionnaire d'évaluation de la pharmacodépendance* [Investigateur principal : Pr Pascale Jolliet (CHU Nantes)]

Appels d'Offre Interne du CHU de Nantes

- Financement 2010 : *Etude multicentrique et prospective évaluant l'impact de l'antibiothérapie sur le microbiote intestinal et la survenue d'une diarrhée.* [Investigateur principal : Dr Emmanuel Montassier (CHU Nantes)]
- Financement 2008 : *Etude monocentrique transversale pour la validation française de deux questionnaires évaluant les distorsions cognitives particulières des joueurs* [Investigateur principal : Pr Venisse (CHU Nantes)]

Financement Association Française contre les Myopathies

- Financement 2007 : *Evaluation de la Qualité de Vie chez les personnes atteintes de dystrophies musculaires : Construction et Validation d'un instrument spécifique* [Investigateur Principal : Dr François Boyer (CHU Reims)]

Validation scientifique (Reviewing)

- Quality of Life Research
- European Journal of Psychological Assessment (EJPA)
- European Review of Applied Psychology (ERAP)
- BMC Health Services Research
- Communications in Statistics - Theory and Methods
- Communications in Statistics - Computation and Simulation
- Mesures et Evaluation

Organisation de manifestations scientifiques

- Membre du comité d'organisation de l'*International Workshop on Patients Reported Outcomes and Quality of Life*, Paris, 4-5 juillet 2011
- Coorganisateur du *Workshop Biostatistique et Mesures Subjectives en Santé*, Nantes, 16 juin 2008.
- Membre du comité d'organisation de l'*International Conference on Statistics in Health Sciences*, Nantes, 23-25 juin 2004

Organisation de formations

- Formation dispensée à la société MAPI-Values (Lyon) sur *la Théorie de Réponse aux Items* à l'équipe de Biostatisticiens, Organisateur de la formation, Financement MAPI-Values, mars 2009
- *Théorie de réponse aux items - Applications sous le logiciel SAS*, Organisateur de la formation, Financement EA 3797 (Reims), décembre 2006
- *Qualité de vie - validation des indices et utilisation dans les essais cliniques*, Co-organisateur de la formation (avec le Pr Mounir Mesbah et le Dr Agnès Hamon), Financement Société Française de Statistique (SFdS), octobre 2002
- *Mesure de la santé perceptuelle et de la qualité de vie. Développements récents et perspectives méthodologiques*, Co-organisateur de la formation (avec les Pr Joël Coste, Alain Leplège et Jacques Pouchot), Ecole doctorale de Santé Publique de l'université Paris Sud - Paris 11, mai 2010

RESPONSABILITÉS PÉDAGOGIQUES

Ecole doctorale Biologie Santé ED 502(Université de Nantes)

- Responsable du séminaire doctoral "Modélisation et Plan d'expérience" (depuis 2008)
- Participation du séminaire doctoral "Biostatistique" (depuis 2008)

Master 2 Recherche Interrégional "Modélisation en Pharmacologie Clinique et Epidémiologie - MPCE" (Universités de Rennes 1, Nantes, Angers, Brest, Tours, Poitiers)

- 2007-2011 : Responsable de l'UE "Données Censurées"
- A partir de 2012 : Responsable de l'UE "Echelles de mesures en Santé"

Master 1 "Bioinformatique et Biostatistique" (Université de Nantes)

- Copporteur (avec le Dr Christine Sinoquet - UFR des Sciences) du projet de création d'un parcours de M1 "Bioinformatique et Biostatistique" dans le cadre du Master Biologie Santé de l'Université de Nantes (ouverture prévue en septembre 2012)
- A partir de 2012 : Responsable de l'UE "Données manquantes et petits échantillons"
- A partir de 2012 : Responsable de l'UC "Introduction à l'Epidémiologie"

UFR de Pharmacie de l'Université de Nantes

- Depuis 2007 : Responsable de l'enseignement de Biostatistique dans le cadre de la préparation au concours de l'internat de Pharmacie (3e, 4e et 5e année de pharmacie)
- Depuis 2011 : Responsable de l'UC "Biomathématiques appliquées à la pharmacocinétique" (2e année de pharmacie)
- Depuis 2011 : Responsable de l'UE "Informatique - Préparation au C2i" (2e année de pharmacie)
- 2006-2010 : TD de "Biomathématiques, Probabilités et Biostatistique" (1ère année de pharmacie)

UFR de Médecine de l'Université de Nantes

- Depuis 2010 : TD de "Biostatistique" (Première Année Commune des Etudes de Santé - PACES)

Ecole Nationale de la Statistique et de l'Analyse de l'Information (ENSAI)

- Depuis 2006 : Cours sur "la Théorie de Réponse aux Items" (3ème année)

RESPONSABILITÉS ADMINISTRATIVES

- Depuis 2011 : Membre du bureau de l'EA 4275
- Depuis 2011 : Correspondant C2i de l'UFR de Pharmacie de Nantes
- Depuis 2009 : Représentant élu des Maîtres de Conférences au conseil de gestion de l'UFR de Pharmacie de l'Université de Nantes
- Depuis 2006 : Président de l'Association pour le Développement de la Biostatistique à Nantes (ADBN)

Chapitre 2

Bilan (1999-2011)

2.1 PARCOURS UNIVERSITAIRE

Mon parcours universitaire a commencé dans le domaine de la Statistique et de l'Informatique, avec un DUT en "Statistique et Traitement Informatique des Données - STID" (Université de Lille II) puis une licence/maîtrise professionnelle en "Génie Informatique et Statistique - GIS" (Université de Bretagne Sud) associées à un titre d'ingénieur-maître dans cette spécialité. Durant ce parcours, pourtant orienté vers la professionnalisation, je me suis découvert un goût pour la recherche et le domaine de la Biostatistique durant des stages réalisés en organisme de recherche ou en milieu hospitalier : Stage sur une "Etude multicentrique sur la pertinence des alarmes en service de réanimation adulte" (laboratoire de Biomathématiques - Université de Lille II) en 2e année de DUT [128], stage sur la "Création de groupes homogènes de malades" (DIM du CH de Vannes) en Licence professionnelle [127], stage sur "Les modèles Linéaires Généralisés : une application pour prédire la présence et l'abondance de quelques mauvaises herbes en Côte d'Or" (laboratoire de malherbologie, INRA Dijon) en Maîtrise Professionnelle [129]. Ceci m'a amené assez naturellement à approfondir ma formation par le DEA en Santé Publique, spécialité Biostatistique à Villejuif/Le Kremlin Bicêtre (Université de Paris 5), où mon stage a porté sur "Capture-Recapture : l'apport du Robust Design en Epidémiologie", réalisé au laboratoire de Statistique Appliquée de Bretagne Sud (Université de Bretagne Sud) [130, 143].

Dans ce dernier stage, j'ai étudié les méthodes de Capture-Recapture appliquée à la santé. Ces méthodes proviennent de la Zoologie où des animaux sauvages sont capturés, bagués, relâchés, puis une seconde vague de capture est réalisée et, à partir de la proportion d'animaux capturés une fois ou deux fois, on peut prédire la taille de la population d'animaux. Ces méthodes sont connues pour avoir une tendance à la sur-estimation des populations, car elles supposent que les populations couvertes lors de chaque capture sont exactement les mêmes, ce qui est rarement respecté (en général, seule un sous-ensemble de la population animale est susceptible d'être capturé à chaque capture). Pour corriger partiellement le biais, il a été proposé, en Zoologie, l'usage du robust design qui consiste à réaliser en plusieurs fois chaque phase de capture afin d'estimer la population couverte par chaque capture. Un modèle adapté permet alors de réduire le biais sur l'estimation de la population générale. Les méthodes de capture-recapture ont été utilisées en Epidémiologie, pour estimer la prévalence de certaines

maladies, en croisant des échantillons de patients issus de différentes listes (médecins généralistes, hôpitaux, associations de malades par exemple). Néanmoins, tout comme en Zoologie, ces listes couvrent rarement exactement les mêmes populations, et donc ces méthodes produisent généralement une surestimation de la prévalence. Nous avons donc proposé d'introduire le Robust Design en réalisant sur chaque liste plusieurs échantillons de patients. Une étude par simulation de Monte-Carlo a montré que ce plan expérimental permettait de réduire considérablement le biais par rapport à la méthode classique, notamment lorsque l'intersection entre les listes (individus couverts par plusieurs listes) était faible. Nous avons aussi montré que le biais résiduel était lié à la "trap-response" (non indépendances des tirages sur les différentes listes) correspondant en pratique au fait que les patients qui était déjà sur une liste avaient moins de chance de se trouver sur une autre liste (processus nommé trap-shyness représentant, par exemple, pour certaines pathologies, le fait qu'un patient suivi par son généraliste a peu de raison d'être aussi suivi à l'hôpital) ou au contraire plus de chance (processus nommé trap-happyness, représentant par exemple le fait qu'un patient adhérent à une association de malade aura plus de chance d'être suivi par un généraliste ou à l'hôpital qu'un autre patient). Le nouvel estimateur avait tendance à surestimer la population en cas de trap-shyness et à la sous-estimer en cas de trap-happyness et à être non biaisé en absence de trap-response.

Suite à ce stage, j'ai entamé un doctorat, qui m'a fait découvrir la spécificité du travail statistique sur les données subjectives issues de questionnaires d'items. Ce champ est devenu mon principal sujet de recherche depuis 1999, et j'ai pu travailler sur différentes phases de l'analyse de ce type de données : la validation des questionnaires, la planification d'essais où le critère principal serait issu de données subjectives, et l'analyse statistique de telles données. J'ai opté pour une présentation dans ce bilan issue de cette classification des travaux, plutôt que sur une présentation purement chronologique.

2.2 VALIDATION D'ÉCHELLE

La validation d'échelle mesurant un concept subjectif fait partie de mes sujets de recherche. Mon sujet de thèse portait sur cette phase particulière de développement d'une échelle, où l'idée est de vérifier la validité et la fiabilité d'une échelle.

Notes techniques. La validation d'échelle

La validation d'échelle de mesure consiste à vérifier si la mesure du concept sous-jacent à cette échelle est une mesure valide et fiable. La validité concerne un certain nombre de propriétés de la mesure, comme la validité de construit (fait que l'échelle mesure bien le concept sensé être mesuré), la validité de face (fait que la personne qui répond comprenne les questions de la même façon que les experts la comprennent, fait que les questions soient compréhensibles et que les modalités de réponses soient exhaustives), la validité de structure (fait que le découpage de l'échelle en dimension soit approprié), la validité concourante (fait que la mesure soit corrélée à des mesures similaires), ou la validité discriminante (fait que la mesure permette de différencier des patients ayant des caractéristiques cliniques différentes). La fiabilité concerne la cohérence interne (fait que le nombre d'items soit suffisant et que les items soient suffisamment corrélés entre eux), la reproductibilité (fait que la mesure bouge peu lorsque l'état

du patient est stable) et la sensibilité au changement (fait que la mesure change lorsque l'état du patient change).

Depuis mon installation nantaise, j'ai eu en outre l'occasion de travailler à la fois sur des aspects méthodologiques, en particulier sur l'impact des données manquantes dans la phase de validation, mais aussi à plusieurs occasions sur des problèmes de validation d'échelles dans différents cadres.

2.2.1 Travaux méthodologiques

Doctorat (1999-2005)

J'ai commencé à travailler sur les données subjectives au cours de ma thèse, débutée en 1999. En effet, mon sujet de thèse, intitulé "Construction d'échelles d'items unidimensionnelles en qualité de vie" [25] avait pour but de comparer les méthodes existantes de sélections d'items en sous-échelles unidimensionnelles à partir d'un ensemble d'items de structure inconnue ou floue, et de définir une méthode permettant d'obtenir des sous-échelles vérifiant un modèle de Rasch. Cette thèse m'a permis d'appréhender la thématique des échelles de mesures multi-items, et notamment la problématique de la validation de ces échelles, ainsi que de me spécialiser dans les modèles de la théorie de réponse aux items (IRT).

Notes techniques. l'IRT et le modèle de Rasch

L'IRT est un ensemble de modèles expliquant les relations entre un trait latent (variable latente) et les réponses aux items d'un questionnaire prévus pour mesurer ce trait latent (variables manifestes). Les modèles de l'IRT sont, pour la plupart d'entre eux, définis de manière paramétrique, mais certains modèles (Guttman, Mokken) sont définis de manière non paramétrique. Les modèles de l'IRT vérifient tous les trois hypothèses fondamentales de l'IRT : unidimensionnalité ou dimension fixée, indépendance locale et monotonie. L'unidimensionnalité (remplacée dans certains modèles par l'hypothèse de dimension fixée) consiste à supposer que les réponses aux items dépendent d'un trait latent unidimensionnel (ou de plusieurs traits latents dont le nombre est connu a priori). L'indépendance locale suppose que toutes les relations entre les items peuvent s'expliquer par la présence du trait latent, ainsi, conditionnellement au trait latent, les réponses aux items sont indépendantes. Enfin, la monotonie induit le fait que plus la valeur du trait latent est élevée, plus la réponse à chaque item est, en moyenne, favorable à ce trait latent, et inversement.

Les modèles de l'IRT offre l'avantage de pouvoir obtenir une mesure d'intervalle du concept latent que l'on cherche à mesurer, c'est-à-dire que des sommes (ou des différences) peuvent être calculées sur la mesure obtenue, et ces sommes (ou ces différences) auront le même sens quelque soit la valeur du trait latent. Cette propriété de la mesure simplifie l'interprétation des différences puisque une même différence s'interprétera toujours de la même manière, quelque soit le niveau de trait latent considéré, et permet de justifier le calcul d'une moyenne sur ce type de mesure (ce qui est utile en Recherche Clinique ou en Epidémiologie ou dès lors que l'on travaille sur des groupes de patients). Elles s'opposent en cela aux scores classiques qui, sur un plan théorique sont habituellement des mesures ordinales (c'est-à-dire qu'elles sont efficaces pour classer les individus mais une différence constante sur un score n'est pas interprétable en

soi, car elle dépendra du niveau de score considéré). Sous certaines conditions très restrictives, un score pourra néanmoins devenir une mesure d'intervalles.

Le modèle de Rasch est le modèle paramétrique le plus connu de l'IRT, prévu pour des items dichotomiques. Dans ce modèle, les items sont caractérisés par une seule propriété : leur difficulté, qui est directement liée à la prévalence des réponses positives à l'item (réponses codées 1, par opposition aux réponses négatives codées 0) : plus les réponses positives sont prévalentes, plus la difficulté de l'item est faible, et inversement. Le modèle de Rasch consiste à modéliser la probabilité d'une réponse positive à un item particulier par modèle logistique où le prédicteur linéaire comprend la valeur du trait latent de l'individu n (θ_n) et l'opposé du paramètre de difficulté de l'item j (δ_j) :

$$P(X_{nj} = x | \theta_n, \delta_j) = \frac{\exp(x(\theta_n - \delta_j))}{1 + \exp(\theta_n - \delta_j)}$$

L'unité statistique est alors la réponse d'un individu à un item, et l'association de l'indépendance locale et de l'indépendance des individus (hypothèse classique de modélisation) permettent de considérer l'indépendance des unités statistiques. On peut noter que la méthode du maximum de vraisemblance classique, appelée en IRT méthode JML (Jointed Maximum Likelihood) est biaisée et qu'il est préférable d'utiliser des méthodes non biaisées, lesquelles dépendent de la nature du trait latent. En effet, soit l'on considère le trait latent comme un ensemble de paramètres fixes (un paramètre par individu) soit comme une variable aléatoire (les manifestations du trait latent observées dans l'échantillon ne sont alors qu'un tirage dans une distribution, et seuls les paramètres de la distribution sont d'intérêt). Dans la modèle de Rasch à effets fixes, les méthodes à privilégier, afin d'obtenir des estimations consistantes des paramètres, sont les méthodes basées sur la maximisation de la vraisemblance conditionnée sur la valeur du trait latent individuel (Conditional Maximum Likelihood (CML) ou Pairwise Conditional Estimation (PCE) en particulier). Dans le cas du modèle de Rasch à effet aléatoire, le maximum de vraisemblance marginale (Marginal Maximum Likelihood (MML)) est la méthode la plus souvent utilisée et permet elle aussi d'obtenir des estimations consistantes des paramètres.

Ma thèse m'a donné l'occasion de développer un nouveau modèle de l'IRT multidimensionnelle, le Multidimensional Marginally Sufficient Rasch Model (MMSRM), qui est un modèle où le nombre de dimensions (traits latents) est connu, et où chaque item est lié à un et un seul trait latent [13]. Ce modèle permet à chaque ensemble d'items (items liés à un trait latent donné) de vérifier un modèle de Rasch et ses propriétés spécifiques : l'exhaustivité du score sur le trait latent et l'objectivité spécifique.

Notes techniques. Les propriétés spécifiques du modèle de Rasch et les modèles de la famille de Rasch

Des données vérifiant un modèle de Rasch vérifient deux propriétés : l'exhaustivité du score sur le trait latent et l'objectivité spécifique. L'exhaustivité du score sur le trait latent est une propriété induisant que pour tous les individus qui ont le même score (somme des réponses à l'ensemble des items), l'estimation du trait latent individuel sera identique. L'objectivité spécifique induit que les estimations des paramètres liés aux items sont indépendantes de l'échantillon étudié, et que les estimations des traits latents individuels sont indépendantes

des items répondus. Par extension, les modèles de l'IRT vérifiant les propriétés d'exhaustivité du score sur le trait latent et d'objectivité spécifique sont appelés "modèles de la famille de Rasch" (One-Parameter Logistic model (OPLM), Partial Credit Model (PCM) et Rating Scale Model (RSM).

A partir du MMSRM, un algorithme a été défini afin de déterminer des ensembles d'items vérifiant chacun un modèle de Rasch dans un ensemble d'items de structure inconnue (c'est-à-dire sans connaissance ni du nombre de traits latents sous-jacents, ni des liaisons items/trait latents). L'idée principale de cet algorithme nommé Raschfit [13] était de sélectionner une paire d'items vérifiant un modèle de Rasch (appelée noyau), puis d'ajouter à ce noyau un nouvel item en comparant via des critères de parcimonie (AIC) deux alternatives : soit un modèle de Rasch avec trois items, soit un MMSRM où le troisième item serait associé à une seconde dimension. Si le modèle de Rasch se révélait le plus parcimonieux, le troisième item était ajouté au noyau, et l'on étudiait l'item suivant, jusqu'à ce qu'aucun nouvel item ne puisse être ajouté au noyau. Cet algorithme était relativement lent (sur les machines de l'époque, il fallait compter environ 5 heures pour un ensemble initial de 14 items et 11 heures pour 21 items). Des optimisations ont été proposées afin d'introduire les items dans un ordre le plus efficace possible. Une seconde version de l'algorithme baptisée Raschfit-fast [28] a été proposée dans le but de réduire le temps de calcul sans perte de qualité dans le classement des items. Le temps de calcul est ainsi passé à environ 5 minutes en moyenne pour 14 items, et à environ 20 minutes pour 21 items, néanmoins, le classement se révélait de moins bonne qualité que Raschfit, notamment lorsque les traits latents étaient corrélés (mais avait de meilleurs résultats que Raschfit pour détecter un ensemble d'items unidimensionnel). Des comparaisons par simulations de Monte Carlo avec d'autres procédures de classement d'items ont été effectuées, notamment avec des procédures issues de l'IRT non paramétriques, cherchant à former des ensembles d'items vérifiant les hypothèses fondamentales de l'IRT (méthodes MSP-AISP, HCA/CCPROX en particulier) mais la comparaison a aussi été réalisée avec des méthodes plus classiques comme l'analyse factorielle exploratoire (AFE).

Notes techniques. Les méthodes MSP-AISP et HCA/CCPROX

MSP (Mokken Scale Procedure), appelée aussi AISP (Automated Item Selection Procedure) est une méthode qui cherche, à partir d'un ensemble d'items de dimension inconnu, à regrouper ensemble les items en fonction de la cohérence entre leur réponse. Cette procédure itérative s'appuie sur le modèle de Mokken (modèle non paramétrique de l'IRT) dans lequel la cohérence entre des items consistent à éviter au maximum des individus qui répondraient positivement à un item difficile, et négativement à un item facile (en sciences de l'éducation, cela suppose de trouver incohérentes les réponses d'un étudiant qui aurait réussi à une question difficile mais échoué à une question facile ; en santé, cela pourrait par exemple concerner un individu qui présenterait dans une pathologie quelconque un symptôme rare, mais ne présenterait pas un symptôme fréquent). MSP-AISP va donc chercher, au sein de l'ensemble d'items, des groupes d'items ayant le moins possible d'incohérence entre leurs réponses. On pourra ainsi associer chaque item à une dimension.

HCA/CCPROX (Hierarchical Cluster Analysis/Conditional proximity measures) est une classification hiérarchique ascendante des items basées sur une distance entre item déterminée à partir des corrélations ou des covariances condi-

tionnelles à la valeur du trait latent des individus. L'idée de cette procédure est de rassembler les items de manière à ce qu'au sein de chaque cluster, l'indépendance locale soit le mieux respectée possible.

Comparé à ces méthodes, l'algorithme Raschfit donnait des résultats corrects en terme de classifications d'items (de l'ordre de ceux de HCA/CCPROX, et meilleurs que MSP-AISP pour une structure simulée multidimensionnelle, et de l'ordre de MSP-AISP, et meilleurs que HCA/CCPROX pour une structure simulée unidimensionnelle), et semblait assez robuste à des écarts à des modèles de Rasch. Le modèle MMSRM et la procédure Raschfit ont été publiés en 2004 [13], et l'algorithme Raschfit-fast a été publié après ma soutenance de thèse, en 2008, dans un chapitre d'ouvrage [28]. Des programmes Stata [111, 113] permettaient d'estimer facilement ce modèle ou de réaliser cette procédure. Ces travaux ont aussi donné lieu à cinq présentations orales dans des congrès internationaux [51, 52, 53, 61, 62]. La procédure Raschfit a été utilisée dans le cadre de la construction d'un score de prédiction du risque de récurrence suicidaire à partir des données cliniques. Ce travail a été publié [26] et présenté dans trois conférences [81, 86, 87].

Ce travail m'a amené en fin de thèse à travailler sur une procédure issue de l'analyse exploratoire multidimensionnelle qui permet de grouper de variables quantitatives au sein d'un ensemble unidimensionnel, la procédure CLV (Clustering around Latent Variables). J'ai testé cette procédure avec des données issues de réponses à des items, donc avec des variables qui, par nature, n'était pas quantitatives, avec des résultats très prometteurs présentés dans une conférence internationale [54]. Un des travaux sur cette procédure a été publié en 2006 [7]. Il consistait à déterminer la structure d'un questionnaire de stigmatisation des patients atteints d'hépatite C. J'ai fait des propositions pour adapter CLV à des données ordinales ou binaires (utilisation des corrélations tétrachoriques ou polychoriques au lieu des coefficients de Pearson, amélioration de l'algorithme), néanmoins, les résultats étaient globalement équivalents, même si sur le plan théorique, cela semblait mieux adapté. Ceci a donné lieu à un programme Stata [112].

Une autre partie importante de ma thèse a consisté à implémenter certaines méthodes ou modèles de l'IRT sous SAS [98-104] et Stata [105-119]. Pour les programmes les plus importants, il s'est agit de programmer l'estimation des paramètres dans un modèle de Rasch, ainsi que des tests et indices d'adéquation. Les modules obtenus (%Anaqol sous SAS, -raschtest- sous Stata) ont été tous les deux publiés en 2007 [8, 9] et présentés dans un colloque international [55]. D'autres modules ont été écrit mais n'ont pas directement donné lieu à des publications (%detect et -detect-, %gausshermite et -gausshermite-, -hcavar- ou -clv- par exemple) ou bien ont été publiés plus tardivement (-loevh- et -msp- sous Stata publiés en 2011 [11]), ou encore ont été présentés au sein d'articles méthodologiques (les modules Stata -mmsrm- et -raschfit- présentés dans [28], les modules Stata -imputeitems-, -imputerasch- et -imputemokken- présentés dans [12], ou le module Stata -raschpower- présenté dans [10]). L'ensemble des programmes est présenté sur un site Internet dédié (www.anaqol.org) et les programmes sont téléchargeables gratuitement.

Ma thèse a été soutenue le 14 novembre 2005 [25], en présence des Pr Catherine Hubert (Présidente), Jean-Louis Auget (Rapporteur), El Mostafa Qannari (Rapporteur), Mounir Mesbah (Directeur de thèse), Michel Chavance (Examinateur) et Paul

Deheuvels (Examineur). J'ai obtenu la mention "très honorable" dans la spécialité "Recherche Clinique, Innovations Technologiques et Santé Publique".

Impact de l'imputation de données manquantes en validation de questionnaire (2007-2011)

Suite à ma thèse, un des premiers travaux méthodologique que j'ai effectué concernait la validation d'échelle, et notamment l'étude de l'impact des données manquantes sur la validation d'une échelle de mesures par un modèle de Rasch, et l'impact des méthodes d'imputation simple habituellement utilisées dans ce cadre. Une étude de simulations de Monte Carlo a été réalisée comparant 14 méthodes d'imputations simples, et deux méthodes d'analyse des données sans imputation des données manquantes (complete case et available case). Les critères étudiés dans cette étude concernaient les biais sur les paramètres classiquement utilisé pour valider une échelle par un modèle de Rasch : paramètres de difficultés des items, variance du trait latent, statistique d'un test d'adéquation, indice de fiabilité des estimations du trait latent (PSI)...

Parmi les méthodes d'imputation, celle qui biaisent le moins les paramètres étudiés étaient globalement les méthodes basées sur une prise en compte simultanée de la difficulté de chaque item, et de la capacité (valeur du trait latent) de chaque individu. En particulier, les méthodes basées sur la vraisemblance du modèle de Rasch faisaient partie des méthodes les moins biaisées. En outre, l'ajout d'un effet aléatoire lors de l'imputation semblait atténuer les biais. Néanmoins, dans tous les cas, l'estimation des paramètres sans imputations en utilisant l'ensemble de l'information disponible (available case) faisait figure de meilleure méthode, mais des biais étaient inévitable dès lors que le taux de données manquantes était grand ($>10\%$). Un autre résultat intéressant montrait un faible impact de l'informativité de données manquantes dans le cadre du modèle de Rasch, un phénomène que nous avons aussi retrouvé dans d'autres travaux sur ce modèle (dans cette étude l'informativité des données manquantes représentait le fait que la probabilité d'observer une donnée manquante pour un individu donné dépendait de la valeur du trait latent de cet individu. Si la probabilité d'observer une données manquantes était indépendante du trait latent de l'individu, alors le processus de génération des données manquantes était non informatif. En outre, en cas d'informativité, plusieurs niveaux d'informativité ont été testé en jouant sur la force de la relation entre la probabilité d'observer une donnée manquante et le trait latent). Cette étude réalisée avec Véronique Sébille a été publiée en 2011 dans BMC Medical Research Methodology [12] et présentée dans un colloque international [63] et deux conférences nationales [84, 85].

2.2.2 Validation d'échelles

Distorsions cognitives des joueurs pathologiques (2007-2011)

Ce premier projet de validation d'échelle concernait le service d'addictologie qui était en train de se spécialiser dans le jeu excessif (pour créer le Centre de Recherche sur le Jeu Excessif - CRJE) et qui voulait valider deux questionnaires de langues anglaises, le GABS (Gambling Attitudes and Beliefs Survey) et le GRCS (Gambling Related Cognition Scale), permettant de mesurer les distorsions cognitives des joueurs, c'est-à-dire les fausses croyances liées au jeu (usage d'objet porte-bonheur, existence de séries bénéfiques ou négatives, croyance en certaines stratégies de jeu sur des jeux de hasard pur...). Aucun outil mesurant ces concepts n'existait à l'époque en langue

française. A l'aide d'un financement interne du CHU de Nantes, une étude a été réalisée comprenant certains joueurs excessifs suivis dans le service, et des étudiants des filières pharmacie et psychologie de Nantes. Pour le GRCS, la structure retrouvée dans la traduction était cohérente avec la structure anglaise en 5 dimensions, et les principales propriétés psychométriques (validité de face, validité concourante, validité discriminante, cohérence interne) ont été retrouvées [4]. Pour le GABS, initialement prévu comme un questionnaire unidimensionnel à 35 items, cette structure n'a pas été jugée cohérente, même si les principaux indices psychométriques (principalement la cohérence interne mesurée par le alpha de Cronbach) étaient cohérents avec la version anglaise. En effet, de meilleures structures en plusieurs dimensions pouvaient être proposées. Ceci a donné lieu à une collaboration importante entre l'EA 4275 et le CRJE pour définir une structure qui permette d'obtenir des dimensions satisfaisantes à la fois sur le plan psychométrique et sur le plan clinique. Une structure en 5 dimensions (stratégie de jeu illusoire, croyance en la chance et les superstitions, contrôle des attitudes de jeu, chasing intra-session (difficulté à arrêter de jouer, dans l'espoir de se refaire), arousal subjectif (activation émotionnelle) lié au jeu) et 23 items a ainsi été proposées. Un article est soumis en ce sens [34]. Ces deux études de validation ont aussi été présentées dans des congrès [92, 95].

Mesure du raisonnement médical (2008-2012)

Un second projet de validation auquel j'ai pu participer concernait l'éducation médicale. En 2007, Pierre Pottier (UFR de Médecine) a créé l'Equipe de Recherche Nantaise en Education Médicale (ERNEM) sur le pôle santé [98]. Un des premiers projets proposé par cette équipe était de travailler sur le raisonnement médical. Parmi plusieurs projets en ce sens, celui de construire une échelle simple de mesure du raisonnement médical a été proposé. La spécificité de ce travail concernait le fait que 4 modes de raisonnement (reconnaissance de scripts, regroupement syndromique, utilisation d'un organigramme, raisonnement hypothético déductif) ont été définis, et il a été rapidement fait constat qu'une seule question pouvait être posée pour chaque mode de raisonnement. Il s'agissait donc de construire une échelle à 4 questions (en fait, une cinquième question a été ajoutée pour différencier deux reconnaissances de scripts différentes : l'une basée sur une connaissance théorique du sujet, l'autre basée sur l'expérience). Dans ce questionnaire particulier, la méthode de validation de l'outil ne pouvait pas faire appel à des méthodes standards (validation de structure, comparaison à une référence, cohérence interne). Nous avons donc préféré une méthode basée sur le remplissage des questionnaires suite à des dossiers cliniques suffisamment typiques pour privilégier l'un ou l'autre des 4 raisonnements. Néanmoins, il a fallu en amont construire ces dossiers et valider le fait que chaque dossier pouvait être associé à un raisonnement clinique particulier.

La première étape a été de trouver des pathologies susceptibles de convenir. Le choix a été fait de construire le dossier basé sur la reconnaissance de script sur une épiglotite, celui sur le regroupement syndromique sur une maladie de Biermer, celui sur l'utilisation d'un organigramme sur un ictère, et celui sur le raisonnement hypothético-déductif sur un prurit. Afin de valider ces dossiers, un protocole basé sur un raisonnement à voix haute d'étudiants et d'experts a été défini : 4 étudiants de 3e année de médecine, 8 étudiants de 4e année de médecine, 8 étudiants de 5e année de médecine, et 4 médecins expérimentés du service de médecine interne du CHU de

Nantes devaient raisonner sur ces 4 dossiers (en aveugle) à voix haute, et leur raisonnement a été analysé par des cartes conceptuelles remplies par deux autres experts initiés à ce type de cartes. L'analyse des 96 cartes ainsi obtenues a permis de montrer des différences importantes de raisonnements entre les experts et les étudiants : plus on acquiert d'expérience, et plus l'on a tendance à utiliser un raisonnement de type inductif (reconnaissance de scripts ou regroupement syndromique) et à valider l'hypothèse formulée à l'aide d'un raisonnement hypothético-déductif. A l'inverse, les plus jeunes étudiants avaient tendance à beaucoup utiliser le raisonnement inductif (en particulier, l'usage d'organigrammes) mais à ne pas valider leurs hypothèses. Néanmoins, les hypothèses sous-jacentes sur les 4 dossiers (utilisation plus ou moins forte des 4 raisonnements dans les 4 dossiers) ont pu être validées.

Aussi dans un deuxième temps, les 4 dossiers et le questionnaire de mesure du raisonnement médical a été posé à 3 promotions consécutives d'étudiants de 5e année de médecine (permettant de cumuler 270 réponses). Les scores de mesure de chaque raisonnement ont été analysés dossier par dossier et ont permis de valider les hypothèses. L'ensemble de ce travail a donné lieu à 2 articles publiés : un premier pour montrer la pertinence des cartes conceptuelles dans ce travail [27] et un second sur la validation des 4 dossiers cliniques [19]. Un article est en cours d'écriture [43] sur la validation du questionnaire de mesure du raisonnement médical, qui a été présenté lors de trois congrès [56] [72] [99].

Mesure de la qualité de vie des patients atteints de malformations anorectales (2010-2012)

Une autre étude de validation de questionnaire à laquelle j'ai été associé concernait la traduction et la validation de plusieurs versions d'un questionnaire de qualité de vie spécifiques aux patients atteints de malformations ano-rectales : le HAQL (Hirschprung and Anorectal Quality of Life). Il existait en effet 3 versions de ce questionnaires en fonction de l'âge du patient : 8-11 ans, 12-16 ans et 17 ans et plus. De plus des versions parents existaient pour les 6-16 ans. Ces 4 questionnaires ont été traduits fin 2009, et les versions enfants (8-11 ans) et parents ont été validées en 2010 par un projet financé par la fondation Pfizer pour la santé des enfants et adolescents (ex fondation Wyeth pour la santé des enfants et adolescents). Melle Pauline Clermidi, interne en chirurgie pédiatrique et étudiante en M2 recherche Modélisation en Pharmacologie Clinique et Epidémiologie a effectué ce travail de validation lors de son stage que j'ai coencadré (à 80%) avec le Pr Guillaume Podevin (service de chirurgie pédiatrique - CHU de Nantes). La structure proposée par les auteurs initiaux (néerlandais) en 9 dimensions et 41 items n'a pas pu être retrouvée, néanmoins, on peut noter que cette structure avait été définie sur les questionnaires adultes. Une autre structure en 7 dimensions et 25 items a été proposée, et un article est en cours de finalisation [35]. De plus, dans le cadre du PHRC MarQol, les questionnaires 12-16 ans (enfants et parents) et 17 ans et plus ont été validés en 2011 par Fanny Feuillet et moi-même. Pour ces deux tranches d'âge, la structure initiale était correcte, néanmoins des petites améliorations à cette structure (groupement de deux fois deux dimensions, suppression d'un item) peuvent être proposées. L'article de validation sera soumis en 2012 [42].

Mesure de la gravité de la Pharmacodépendance (2008-2012)

Dans un autre contexte, je suis associé à un PHRC visant à valider un questionnaire de mesure de la gravité de la pharmacodépendance. L'échelle utilisée est tirée de l'échelle de mesure de la dépendance aux substances en 7 items, à laquelle un 8e item sur le comportement de transgression a été ajouté. Cette échelle est déjà utilisée en routine au Centre d'Investigation et d'Évaluation de la Pharmacodépendance (CEIP) du CHU de Nantes, mais n'avait pas été validée. L'analyse porte sur 3180 dossiers, dont 672 dossiers complets. Les premiers résultats montrent une structure plutôt de type bidimensionnel avec une distinction de deux groupes d'items : d'une part les signes physiques et compulsifs de la dépendance (tolérance, symptômes de sevrage, augmentation des doses, désir d'arrêt), d'autre part les conséquences dommageables de la dépendance (temps passé, problèmes relationnels, problèmes de santé, comportement de transgression). La validation complète doit avoir lieu en 2012.

Construction d'un questionnaire de qualité de vie pour les patients atteints de maladies neuromusculaires (2007-2012)

En 2007, j'ai été associé à la construction d'un questionnaire de qualité de vie en langue française pour les patients atteints de maladies neuromusculaires, en association avec l'EA3797 du CHU de Reims (Direction Pr François Boyer). L'absence de questionnaire spécifique de qualité de vie dans ces pathologies a nécessité le développement complet de l'outil (définition des concepts, des items, sélection des items, passation auprès d'un échantillon de patients d'un outil important (112 items), resélection des items). La version pré-validation est actuellement soumise à un échantillon de patients, et je serai responsable de l'analyse statistique pour cette validation, sachant que cette analyse devra se faire avec les méthodes classiques ainsi qu'avec de l'IRT. La démarche de ce projet de recherche a été présentée dans des conférences nationales [78, 79, 80].

2.3 PLANIFICATION D'ÉTUDES PORTANT SUR DES MESURES SUBJECTIVES

2.3.1 Comparaison des approches CTT et IRT : de la comparaison de groupes au projet PLANIPRO (depuis 2007)

En 2007, avec Véronique Sébille, nous nous sommes lancés dans la comparaison des deux approches (CTT et IRT) pour comparer deux groupes de patients dans un essai classique (non séquentiel), avec ou sans données manquantes, toujours par le biais de simulations de Monte Carlo. Rapidement, les résultats trouvés nous ont déroutés, car les puissances que nous obtenions avec les deux approches étaient très proches en l'absence de données manquantes, mais surtout, cette puissance était moindre que celle que nous nous attendions à trouver, estimée par les formules de détermination de nombre de sujets nécessaires classiques (en les inversant pour estimer la puissance en fonction du nombre de sujets simulés).

En effet, si le fait que l'approche CTT était moins puissante que prévu ne nous semblait pas incohérent, étant donné la nature ordinaire du score, et le fait que le calcul de la puissance était réalisée pour des variables continues, nous ne comprenions pas pourquoi la comparaison réalisée avec un modèle IRT n'avait pas la puissance

attendue, étant donné que cette comparaison était réalisée par un test de Student sur les moyennes des traits latents estimées dans le cadre de ce modèle, et que, dans la formule de la puissance utilisée, nous donnions bien des paramètres relatifs à ce trait latent (principalement la différence de moyenne attendue et la variance du trait latent). Ce point a été creusé, ultérieurement pour aboutir à une compréhension de ce phénomène, expliqué par deux raisons différentes selon la méthode :

- Pour la CTT, la taille d'effet (effect size) mesurée par la différence entre les moyennes des deux groupes divisée par l'écart-type de la variable d'intérêt n'a pas la même valeur que pour l'IRT. En effet, en CTT, on travaille sur les scores alors qu'en IRT, on travaille sur la variable latente, et bien que corrélées, les deux tailles d'effets diffèrent. Ainsi, la taille d'effet mesurée sur le score est plus faible que celle mesurée sur le trait latent, ce qui aboutit mécaniquement à une moindre puissance en CTT
- Pour l'IRT, le trait latent est par nature une variable latente, qu'il faut avant tout estimer à l'aide du modèle de mesure, le modèle de Rasch par exemple, ce qui aboutit à une incertitude sur les valeurs de cette variable, laquelle incertitude n'est pas prise en compte dans l'estimation classique de la puissance. En conséquence de quoi, il est légitime de considérer que la formule classique de détermination du nombre de sujets nécessaires, et donc par extrapolation la formule d'estimation de la puissance, seront inadéquates pour travailler sur des variables latentes, puisqu'elles sont prévues pour des variables manifestes (observées) ne nécessitant pas d'estimation au préalable

Ce constat a été publié dans *BMC Medical Research Methodology* en 2010 [22] et présenté dans différents colloques [58, 59, 65, 66, 67] [73] [94]. Il est à l'origine du projet PLANIPRO visant à proposer des estimations adéquates du nombre de sujets nécessaires dans le cadre des modèles IRT, notamment dans l'objectif de comparer des groupes de patients. Ce projet a été financé par l'ANR en 2009 pour un projet prévu sur la période 2011-2013, projet dans lequel je suis coporteur.

Dans le cadre de ce travail, une méthode a été proposée pour déterminer le nombre de sujets nécessaires dans le cadre de la comparaison de groupe sur un trait latent mesuré par un modèle de Rasch. L'idée était d'évaluer dans un premier temps la puissance que l'on pouvait espérer dans un test de Wald testant la nullité d'un paramètre de groupe inclus dans un modèle de Rasch (le trait latent est alors décomposé en une partie fixe expliquée par l'appartenance à un groupe ou à l'autre, et une partie aléatoire représentant le trait latent résiduel, c'est-à-dire des disparités individuelles au sein de chaque groupe). Lors d'un stage, Sarah Amri (ENS Cachan, antenne de Bretagne) a développé analytiquement la statistique de test. L'idée était alors de chercher une formule analytique permettant d'exprimer la puissance du test en fonction des différents paramètres (nombre d'individus dans les deux groupes, différence de moyenne sur le trait latent entre les deux groupes, variance du trait latent, difficultés des items, risque de 1ère espèce). Néanmoins, rapidement, le constat a été fait que l'écriture analytique de la puissance n'était pas possible. En effet, la statistique du test de Wald nécessite d'estimer l'erreur standard du paramètre de groupe, or pour obtenir celle-ci, il est nécessaire de connaître les réponses des individus aux items (ce qui est bien évidemment inconnu en phase de planification d'une étude) et en plus, l'approximation de cette erreur standard ne peut se faire que numériquement (et non analytiquement) car son écriture dépend d'intégrales impossibles à calculer de façon analytique.

Une solution a été proposée pour obtenir des réponses des individus aux items dans le cadre de la planification via un algorithme qui permet de créer un fichier de données de toute pièce en se basant sur les hypothèses faites dans cette phase de planification. Cette base de données, appelée matrice de planification, correspondait à un jeu de données théorique vérifiant parfaitement les hypothèses faites, et correspondant au jeu de données attendu sous ces hypothèses. Pour le créer, tous les profils de réponses possibles sont étudiés (dans chaque groupe de patients) et pour chacun d’eux, on estime la probabilité qu’il survienne selon le modèle supposé (dans le cadre de ce projet, un modèle de Rasch avec une covariable représentant le groupe). A partir de ces probabilités, pour chaque profil de réponse possible, on affecte un effectif en fonction des nombres de sujets dans chaque groupe. L’association de chaque profil de réponses à un effectif permet ainsi de définir la matrice de planification, à partir de laquelle on va estimer l’erreur standard du paramètre représentant le groupe. A partir de ceci, la puissance peut être estimée par la formule

$$1 - \beta = 1 - \Phi \left(z_{1-\alpha/2} - \frac{\gamma}{s.\hat{e}.(\hat{\gamma})} \right)$$

avec Φ la fonction de répartition d’une loi normale centrée réduite, $z_{1-\alpha/2}$ le $1-\alpha/2$ ème percentile de la loi normale centrée réduite, γ l’effet groupe attendu et $s.\hat{e}.(\hat{\gamma})$ l’estimation de l’erreur standard du paramètre γ . Cette approche a été validée à l’aide de simulations de Monte Carlo. Elle permet de donner une estimation correcte de la puissance à 0.02 près. Cette méthode a été publiée dans *Statistics in Medicine* en 2012 [10] et présentée dans des congrès internationaux [67, 66]. Cette méthode est actuellement en cours de validation [32] pour vérifier son adéquation dans des cas où le trait latent serait très hétérogène (variance très grande) ou lorsque les items seraient mal adaptés à la population étudiée (difficultés très petites ou très grandes). Des extensions sont en cours d’élaboration pour le cadre polytomique [57] [41]. Cela a nécessité de proposer des approches permettant de définir rapidement les profils de réponses les plus pertinents (les plus probables), car le nombre de profils de réponses possibles devient vite gigantesque dans le cadre polytomique ce qui pose des problèmes informatiques.

Un résultat surprenant de ce travail a été de montrer que le ratio N/N_C avec N le nombre d’individus utilisés dans l’approche précédente et N_C le nombre d’individus nécessaire pour obtenir la même puissance que celle trouvée avec ces N individus et la formule classique (pour variables manifestes) [laquelle est inadaptée et tend à surestimer la puissance] a une valeur supérieure à 1 et est très stable, quelle que soit la valeur de N . Cette stabilité permet donc, à partir du nombre de sujets nécessaires obtenus avec la formule pour variable manifeste pour une puissance donnée, de retrouver une taille d’échantillon correcte pour analyser une variable latente. L’étude de ce ratio est en cours pour déterminer s’il est possible de l’estimer rapidement par une formule analytique [47].

2.3.2 Nombre de sujets nécessaires pour la validation de questionnaire (Depuis 2011)

Un autre domaine dans lequel il est difficile de déterminer de façon satisfaisante un nombre de sujets nécessaires concerne la validation de questionnaires. La littérature donne des valeurs indicatives contradictoires et rarement justifiées sur le plan scientifique. On trouve par exemple des ordres d’idées allant de 200 à 1000 individus, ou des formules obtenues de façon empiriques de type $5K$ ou $10K$ (où K représente le nombre

de scores différents pouvant potentiellement être obtenus (par exemple, pour valider une échelle à 10 items ayant 3 modalités de réponses (31 scores possibles), il faudrait entre 155 et 310 individus avec ces formules). Ces formules sont peu satisfaisantes sur le plan scientifique, voire peuvent poser question (étant donné que plus une échelle a d'items et plus la mesure sera généralement fiable, donc on pourrait même penser intuitivement que le nombre de sujets nécessaires diminuerait avec le nombre d'items).

Dans ce cadre, Emmanuelle Anthoine a commencé une thèse en 2011, dont la première partie consistera à faire une revue de la littérature de tous les articles de validations d'échelles publiées au niveau international entre 2009 et 2011. L'idée de cette revue de la littérature est d'avoir une idée des nombres de sujets utilisés dans les études de validation d'échelles (que ce nombre soit justifié ou non) et d'avoir une idée des formules (lorsqu'il y a justification) pour déterminer ce nombre de sujets. En parallèle, une revue des méthodes employées pour la validation des échelles sera réalisée.

Les validations de traductions dans une autre langue que la langue originelle de l'échelle, ou les propositions de modification d'échelles existantes ont été exclues de cette revue de la littérature. 4541 articles ont été trouvés dans la littérature, et après lecture des titres par deux lecteurs indépendants (Emmanuelle Anthoine pour l'ensemble des titres, Leila Moret, Véronique Sébille et moi-même pour un tiers des titres chacun), 1046 articles ont été sélectionnés (23%). Les discordances ont été arbitrées par un troisième juge. Une seconde étape s'est déroulée de la même manière à la lecture des 1046 résumés et a permis de sélectionner 422 articles (40%). 130 articles (31%) tirés au sort sont actuellement en cours d'analyse.

2.4 ANALYSE DE DONNÉES SUBJECTIVES

2.4.1 Travaux méthodologiques

Analyses séquentielles sur données subjectives (2004-2008)

Avant la fin de mon doctorat, j'ai entamé une collaboration avec Véronique Sébille, afin de comparer par simulations de Monte Carlo, deux manières d'appréhender des données subjectives multi-items (méthode classique utilisant le score (CTT), méthode impliquant un modèle IRT) dans le cadre des essais séquentiels (études incluant des règles d'arrêt précoce, basée sur un recrutement de petits groupes de patients, avec une analyse statistique répétée à la fin de chaque inclusion d'un groupe de patients, laissant la possibilité de conclure précocement dès lors que le risque de première espèce est globalement maîtrisé, et la puissance suffisante) lorsque le critère principal est une mesure subjective, et en présence de données manquantes, informatives ou non (c'est-à-dire dépendantes du trait latent des individus ou non). Cette collaboration a montré un avantage certain à utiliser un modèle IRT par rapport à la méthode classique (CTT) lorsque les données manquantes étaient nombreuses, et ce quelle que soit l'informativité des données manquantes : la puissance était plus importante, et le nombre de sujets inclus moins important, tout en maintenant correctement le risque de 1ère espèce. Ce travail a été publié en 2007 dans *Statistics in Medicine* [23] et a été complété par une comparaison de ces deux méthodes dans le même contexte mais en l'absence de données manquantes, publié en 2008 sous forme de chapitre d'ouvrage

[29] et présenté dans un colloque international [68].

Cette première collaboration a permis de préciser des résultats retrouvés et expliqués par la suite, en particulier que sans données manquantes, les deux approches CTT et IRT avaient des puissances comparables, et que l'IRT se comportait mieux que la CTT en présence de données manquantes, avec un impact très faible de l'informativité de celles-ci. Ce travail a été achevé lors de mon recrutement comme ATER dans le service de Biomathématiques et Biostatistique de la faculté de pharmacie de Nantes, au moment du dépôt de dossier pour demander la création de l'équipe EA4275 par Véronique Sébille. Cette collaboration a permis de lancer les grandes idées de la thématique Biostatistique et Mesures Subjectives en Santé : continuer de comparer les deux approches CTT et IRT, dans le cadre des essais séquentiels, mais aussi dans le cadre des essais plus classiques, mesurer l'impact des données manquantes et rechercher la meilleure façon de tenir compte de ces données manquantes.

Comparaison des méthodes CTT et IRT pour l'analyse de données transversales (depuis 2010)

La comparaison de deux groupes de patients dans le cadre des modèles IRT a été rapidement un des sujets de recherche de l'EA 4275, et c'est ce point particulier qui a permis de mettre en évidence l'inadéquation de la formule de détermination du nombre de sujets nécessaires lorsque la variable d'intérêt était une variable latente, ce qui a abouti au projet PLANIPRO.

Ce dernier projet profite en parallèle de la thèse de Jean-François Hamel (depuis 2010) dans laquelle je suis coencadrant (à hauteur de 60%). Jean-François Hamel travaille sur les méthodes de comparaison de groupes dans les modèles IRT. Il a notamment comparé 12 méthodes différentes de comparaison de groupes (CTT + 11 méthodes basées sur l'IRT), en fonction de la nature du trait latent (effets fixes ou effet aléatoire), du niveau auquel est faite la comparaison (directement au sein du modèle de mesure, ou ultérieurement à partir des estimations individuelles du trait latent) et de la manière d'estimer les traits latents individuels (valeur modale, imputation simple, imputation multiple) à l'aide de simulations de Monte Carlo. Dans ce cadre, il a été montré notamment que la méthode la moins biaisée et la plus puissante en IRT et en absence de données manquantes pour comparer deux groupes consistait à tester l'ajout d'une covariable groupe directement dans le modèle à l'aide d'un test de Wald. Cette méthode donnait des résultats comparables à l'analyse basée sur la CTT. L'article présentant ces résultats est en cours de révision [30] et présentés sous forme de poster dans une conférence internationale [71]. Néanmoins, en présence de données manquantes, la méthode basée sur la CTT devient moins performante que celle basée sur l'IRT (effondrement de la puissance si l'on n'impute pas les données manquantes et maintien de la puissance mais apparition de biais sur les paramètres en imputant les données manquantes) alors que la comparaison basée sur l'IRT garde de bonnes propriétés (absence de biais, maintien de la puissance), et ce quelque soit l'informativité des données manquantes. Les premiers résultats ont été présentés dans une communication orale [50] et un article est soumis [39]. Le travail se poursuit afin d'étendre ces analyses aux items polytomiques et de mesurer la robustesse de ces méthodes de comparaisons de groupes, notamment à un non respect des hypothèses du modèle de Rasch à effet aléatoire (normalité du trait latent, indépendance locale,

absence de fonctionnement différentiel des items...).

Comparaison des méthodes CTT et IRT pour l'analyse de données longitudinales (depuis 2008)

Début 2008, nous avons bénéficié d'une bourse de la ligue contre le cancer pour financer la thèse de Myriam Blanchin, que nous avons coencadrée, Véronique Sébille et moi-même (à hauteur de 50%). Cette thèse portait sur la comparaison des approches CTT et IRT dans le cadre des analyses de données subjectives longitudinales. Un premier article a montré la nécessité en IRT d'écarter les procédures en deux étapes (une première étape servant à estimer les traits latents individuels et une seconde à estimer les paramètres de la tendance longitudinale (moyennes à chaque temps et matrice de covariance entre les mesures)). Il vaut mieux privilégier une approche en un seul modèle permettant de mesurer le trait latent, et d'estimer simultanément les paramètres de la tendance longitudinale. Dans ce premier article, les comparaisons entre les méthodes étaient effectuées par des simulations de Monte Carlo, sans aucune donnée manquante. Dans ce contexte encore, les deux méthodes CTT et IRT donnaient des résultats comparables. L'article a été publié dans *Statistics in Medicine* en 2011 [1] et présenté dans quatre conférences [48] [77] [89, 90]. Un second article publié dans *IJAMAS* en 2011 [2] (résultats présentés précédemment dans deux conférences internationales [49] [69]) incluait des sorties d'études prématurées (dropout), informatives ou non (c'est-à-dire liées ou non au trait latent du patient, en supposant par exemple qu'un patient qui supporterait mal un traitement et verrait sa qualité de vie chuter, aurait plus de chance de quitter prématurément une étude). Contrairement à d'autres contextes présentant des données manquantes (essais séquentiels ou comparaison de groupe dans une étude transversale), les deux méthodes (CTT et IRT) montraient des puissances similaires. Il était néanmoins à noter que le dropout impliquait la présence ou l'absence complète d'un questionnaire à un temps précis, alors que dans les autres contextes, cela ne concernait qu'une partie des réponses aux items. Dans un contexte d'absence globale du questionnaire, l'IRT ne pouvait pas mieux estimer le concept à mesurer que la CTT.

Un dernier article lié à cette thèse, encore en phase de soumission [33], ajoutait un effet groupe dans les études longitudinales avec dropout informatif ou non. L'analyse de cet effet groupe montrait que le risque de première espèce était bien maîtrisé en l'absence de dropout, mais augmentait en présence de dropout (pour l'analyse CTT ou l'analyse IRT). La puissance liée au test diminuait lorsque le taux de dropout augmentait, et lorsque la corrélation entre les temps augmentait (avec les deux approches). Enfin, quelle que soit l'approche utilisée, l'effet groupe était estimé sans biais.

La thèse de Myriam Blanchin a été soutenue en juin 2011. Grâce à une allocation de recherche, le travail entamé a pu être poursuivi à partir de janvier 2011 par Elodie De Bock-Dumas que je coencadre aussi à hauteur de 60%. Elodie De Bock-Dumas s'intéresse au même contexte (données longitudinales portant sur l'analyse d'une échelle de mesure), avec une comparaison des deux approches CTT et IRT à l'aide de simulations de Monte Carlo, mais introduit des données manquantes partielles sur les questionnaires à chaque temps. Dans une première étude de simulation, elle a montré des résultats très cohérents avec ceux obtenus dans les analyses transversales : sans imputer les données manquantes, la puissance en IRT se maintient bien alors

que cette de l'approche CTT s'effondre, mais en imputant les données manquantes en CTT, des biais sur les paramètres apparaissent. En outre, l'informativité des données manquantes semblent avoir peu d'impact sur les performances de l'approche IRT. Ces premiers résultats ont été présentés sous forme d'un poster dans un colloque international [70] et un article est en cours d'écriture [37].

Estimation des paramètres du modèle de Rasch (2009-2011)

En parallèle à ce travail méthodologique, un travail technique a été tenté ces dernières années en collaboration avec Mohand-Larbi Feddag et Véronique Sébille pour chercher à définir une méthode d'estimation plus rapide pour les modèles de la famille de Rasch. Il s'agit de la Pairwise Likelihood Estimation, qui a donné lieu à un article [6], mais qui, hélas, n'a pas permis d'obtenir des estimations efficaces plus rapidement que la méthode standard (Marginal Maximum Likelihood - MML).

L'analyse en classes latentes (Depuis 2009)

En 2009, Fanny Feuillet a réalisé un stage de Master 2 sur l'analyse de la pharmacodépendance de patients prenant des psychotropes. Les données étaient issues des fichiers de remboursement à l'échelon régional du service médicale de la Caisse nationale de l'assurance maladie des travailleurs salariés (CNAMTS). L'objectif de ce stage était de caractériser ces patients en terme de pharmacodépendance en se basant sur les manifestations visibles de cette dépendance (nomadisme médical ou pharmaceutique, augmentation des doses, ...). Une telle classification représente un concept latent qualitatif, puisque la pharmacodépendance ne peut être qu'indirectement mesurée qu'à travers de telles manifestations

Ce travail de stage est décrit à la section 2.4.2, mais suite à celui-ci, nous avons cherché à comparer deux approches pour créer des groupes de patients à partir de ces données : d'une part l'Analyse en Classes Latentes (LCA) et d'autre part, une Classification Ascendante Hiérarchique (CAH) précédée par une Analyse des Correspondances Multiples (ACM) afin de transformer les données binaires en axes factoriels quantitatifs. Les résultats montraient une certaine cohérence entre les deux approches, sur le nombre de groupes formés (4), et les proportions d'individus dans chaque groupe. Néanmoins, en analysant de près les divergences entre ces deux méthodes, on note que l'approche CAH avait tendance à regrouper ensemble des profils homogènes sur les réponses (ce qui pouvait être attendu, étant donné que la CAH regroupe les profils de réponses en se basant sur leurs distances géométriques), ce qui était moins le cas pour la LCA (qui elle, cherche plutôt à regrouper des profils de manière à expliquer, au sein de chaque groupe, les relations entre les variables étudiées, par la seule appartenance au groupe (principe d'indépendance locale)). Les divergences portaient néanmoins sur des profils de réponses peu fréquents, et l'interprétation pharmacologique des groupes a été plus aisée à partir des résultats de la LCA. Un article est en cours de rédaction pour présenter ces résultats [38]. Fanny Feuillet vient d'ailleurs d'entamer une thèse sur l'apport de l'approche LCA pour l'analyse de données de pharmacodépendance.

2.4.2 Travaux appliqués

Analyse longitudinale de la qualité de vie et du coping des patientes atteintes d'un cancer du sein et de leur accompagnant (2007-2009)

Dans le cadre du projet SALTO (Suivi A Long Terme en Oncologie), un axe de recherche portait sur l'analyse conjointe de la qualité de vie et du coping de patientes atteintes d'un cancer du sein, en liaison avec la qualité de vie et le coping de leur accompagnant, dans un cadre longitudinal. Les patientes (et leur accompagnant) devaient répondre à un questionnaire de qualité de vie (QLQ-C30 pour les patients, DHP pour les accompagnants) et à un questionnaire de coping (WCC) à 3 temps de l'étude : après le diagnostic, après la fin des traitements et à 6 mois après la fin des traitements. Dans le cadre du stage de Master 2 de Julie Léger (que je coencadrait à hauteur de 50%), une typologie des patientes en 4 classes a été proposée, basée sur une Classification Hiérarchique, et les données ont été analysées par un modèle mixte intégrant le caractère longitudinal des données (les scores de qualité de vie des patientes étaient analysés en fonction de la classe d'appartenance des patientes, de ses scores de coping, ainsi qu'en fonction des scores de qualité de vie et de coping de l'accompagnant). L'analyse montrait des influences des niveaux de qualité de vie et de coping de l'accompagnant sur l'évolution de la qualité de vie des patientes et montrait l'importance de tenir compte de l'accompagnement dans la prise en charge médicale et psychologique de ces femmes. Ces travaux ont été publiés en 2012 [3] et présentés dans deux conférences [88] [96].

Analyse spatio temporelle de la qualité de vie en population générale (Depuis 2009)

Depuis 2009, je travaille avec certains membres de l'équipe APEMAC (EA 4360) [Joël Coste, Alain Leplège, Etienne Audureau] sur l'analyse par un modèle de Rasch de deux grands échantillons représentatifs de la population française réalisés en 1995 (par la SOFRES) et 2003 (Enquête décennale INSEE) auprès desquels le questionnaire SF36 a été posé. L'idée de cette analyse est d'utiliser le modèle de Rasch afin d'obtenir une mesure ayant des propriétés de mesure d'intervalle, et invariante entre les deux passations, c'est-à-dire indépendante de la perception des items par la population au sein de chaque échantillon. Dans cette analyse, nous avons tout d'abord recherché du Fonctionnement Différentiel des Items (DIF) entre les deux sexes, les différentes tranches d'âge, les deux échantillons ou entre les régions de domicile. Aucun DIF n'a été détecté entre les deux échantillons ou entre les régions, par contre du DIF a été détecté entre les sexes ou entre certaines classes d'âge. Celui-ci a été pris en compte pour analyser l'évolution de la qualité de vie mesurée par le SF36 entre les deux périodes. Dans cette analyse, on retrouve les effets connus entre les tranches d'âges (diminution de la qualité de vie avec l'âge), ou entre les sexes (qualité de vie moindre chez les femmes), mais aussi on note des différences sensibles entre les régions françaises et une diminution significative de la qualité de vie déclarée entre les deux échantillons. Ces résultats ont été aussi analysés avec les méthodes classiques (analyse du score) [76], qui ne pouvait pas prendre en compte le DIF. Ces deux analyses retrouvaient les mêmes tendances fortes (impact de l'âge, du sexe, différences entre certaines régions, diminution de la qualité de vie entre les deux échantillons). Néanmoins, l'approche par le modèle de Rasch permettait de mettre en évidence de plus nombreuses différences entre les régions, ou interactions entre les variables. Ces différences peuvent s'expli-

quer par la prise en compte du DIF en IRT, mais aussi par la présence de données manquantes de manière non négligeable (13% des individus étaient concernées et 7% présentaient plus de 50% de données manquantes). En outre les données manquantes semblaient très informatives. Dans un tel contexte, l'IRT reste la méthode qui présente la meilleure puissance (et le moins de biais sur les paramètres). Les résultats issus du modèle de Rasch ont été présentés dans une conférence internationale [60] et un article est soumis [40].

Education Médicale (Depuis 2007)

Mon implication dans l'équipe de recherche nantaise en éducation médicale (ERNEM) s'est traduite aussi par l'analyse de données issues de deux expériences auprès des étudiants de 5e année de Médecine.

Dans une première expérience, des étudiants de 5e année étaient placés dans deux conditions expérimentales, où ils devaient définir pour un patient le diagnostic et la prise en charge adéquate : dans un premier temps auprès d'un patient hospitalisé, et dans un second temps, auprès d'un patient lors d'une consultation en ambulatoire. Avant et après chaque visite, le stress de l'étudiant était mesuré par des échelles de stress (stress psychologique) et par un dosage de cortisol salivaire (stress biologique). L'hypothèse de cette étude était que les étudiants seraient plus stressés par une consultation ambulatoire car, d'une part, ils pratiquent peu (voire pas du tout) cet exercice durant leurs études, et d'autre part, parce que les attentes du patient, et la responsabilité individuelle du médecin sont plus fortes en consultation qu'en milieu hospitalier. Les résultats de cette étude montraient un stress plus élevé en situation de consultation ambulatoire, en particulier pour les filles sur les mesures psychologiques réalisées avant la consultation et chez les garçons pour le stress biologique. Les résultats de cette expérience ont été publiés dans *Medical Education* [18] et présentés dans deux conférences internationales [64] [75].

Dans une seconde expérience, les étudiants recevaient en consultations deux acteurs jouant le rôle d'un patient et de son accompagnant. Dans un premier temps d'apprentissage, les acteurs étaient conciliants. Dans un second temps, les étudiants ont été randomisés en deux groupes : d'une part un groupe avec des acteurs conciliants, d'autre part un groupe avec des acteurs créant une atmosphère de stress intense (remise en cause de la parole de l'étudiant, malaise du patient, impatience...). L'hypothèse était de montrer qu'en situation de stress, le raisonnement médical pourrait être moins efficace. Le stress de l'étudiant était mesuré par des échelles à la sortie de la consultation et par la mesure du taux de cortisol salivaire. Nous n'avons pas pu mettre en évidence de liens entre stress et détérioration du raisonnement médical, étant donné que nous avons observé peu de différences en terme de stress mesurable entre les deux groupes d'étudiants, ces derniers semblant plus stressés par l'expérience (première expérience pour ces étudiants de consultations seuls, même s'ils savaient qu'ils avaient affaire à des acteurs), que par le comportement des acteurs. Par contre, le comportement et la communication des étudiants (mesurés par les acteurs sur des échelles spécifiques) semblaient plus dégradés dans le groupe avec des acteurs stressants. Les résultats de cette étude ont été soumis à une revue internationale [45].

Cette étude du lien entre le stress et le raisonnement médical se poursuit actuellement au sein de l'ERNEM. Début 2012, cette étude a été reconduite auprès des étudiants de troisième année de Médecine de Nantes, avec une amélioration du design d'étude : d'une part, les étudiants avaient deux séances d'entraînement de consul-

tation avec des acteurs manière à diminuer le stress lié à la nouveauté, et dans un second temps, chaque étudiant a été pris comme son propre témoin dans un design de type cross-over où chaque étudiant allait successivement à deux consultations, une en situation non stressante, et une en situation stressante (avec tirage au sort de l'ordre des consultations). Les marqueurs de stress biologiques et psychologiques ont été relevés avant et après chaque consultations, ainsi que le raisonnement médical et le comportement de l'étudiant. Les données sont actuellement en cours de saisie.

Pharmacodépendance (Depuis 2007)

Enfin, d'autres travaux ont porté sur la pharmacodépendance, en collaboration avec les collègues pharmacologues de l'EA 4275. Tout d'abord, dans le cadre d'une étude pour caractériser la pharmacodépendance aux benzodiazépines (BZD) chez les sujets âgés, à partir de 184 questionnaires de mesures de la gravité de la pharmacodépendance répondus par des pharmaciens d'officine pour des sujets âgés de plus de 65 ans vis-à-vis de leur usage des BZD. Cette étude montrait une forte prévalence de la dépendance aux BZD des sujets âgés (environ 35%), avec souvent des signes de sevrage (52%), du temps passé pour obtenir leur prescription (47%), une tolérance aux traitements (35%), et une incapacité à arrêter leur traitement (31%). Cette étude a été publiée en 2010[14].

Dans un registre différent, nous avons travaillé sur la caractérisation de la consommation de psychotropes à l'aide d'une base de données de l'échelon régional Pays de la Loire du service médical de la Caisse Nationale d'Assurance Maladie des Travailleurs Salariés (CNAMTS) incluant toutes les délivrances de psychotropes sur le premier semestre 2008. Nous avons ainsi caractérisé les consommations de trois psychotropes : Bromazepam, Zolpidem et Paroxetine en utilisant des analyses en classes latentes. Pour le Bromazepam et le Zolpidem, 4 groupes de patients ont été identifiés : les utilisateurs non problématiques, les patients à risque, les patient avec un trouble mental probable, et les utilisateurs compulsifs. Pour la Paroxetine, seuls trois groupes ont été identifiés, qui se différenciaient plutôt sur le profil de prescription. Cette analyse a été réalisées durant les stages de Master 2 de Fanny Feuillet et Laura Wainstein, et a montré l'utilité de l'analyse en classes latentes dans ce contexte. Un article a été publiée en 2011 [24] et ce travail a aussi été présenté dans une conférence [93].

Analyse du vécu des étudiants en Première Année Commune des Etudes de Santé à Nantes (2011)

En 2011, avec la mise en place de la Première Année Commune des Etudes de Santé (PACES), une étude a été lancée sur Nantes par le département SHS de l'UFR de Médecine (Angélique Bonnaud-Antignac) et a été relayée par les établissements concernés par cette année (UFR de médecine, pharmacie et odontologie, école de sage-femmes) afin d'évaluer les conditions de vie, le stress et la qualité de vie des étudiants de PACES de Nantes. Deux fois dans l'année (15 jours après la rentrée, puis en janvier après les premiers résultats), les étudiants ont été invités à remplir des questionnaires. L'analyse statistique a été réalisée par Mathilde Queré que j'ai encadrée pour son stage de DUT STID.

Le taux de retour était correct (supérieur à 70%). Le stress des étudiants semblait important (plus élevé que celui des premières années toutes filières confondues,

en comparaison à une étude réalisée à l'université de Bordeaux II sur l'ensemble des étudiants de première année) et la qualité vie mesurée par le SF36 était sensiblement dégradée (largement en dessous du niveau des 18-25 ans et au niveau d'une population de patients faiblement ou modérément déprimés). Les étudiants présentaient en outre des taux importants de symptômes anxieux (72% avaient un score de 8 ou plus au score Anxiété du HADS, signifiant un trouble anxieux probable) ou dépressifs (42% avaient un score de 8 ou plus au score Dépression du HADS, signifiant un trouble dépressif probable). Une typologie des étudiants a en outre été réalisée, permettant de définir 5 groupes d'étudiants aux comportements distincts. Ces résultats ont été présentés à une conférence nationale [91].

Il est prévu de renouveler cette étude au cours des prochaines années sur les étudiants de PACES, mais aussi de suivre des cohortes d'étudiants au cours de leur cursus (pour les étudiants admis en deuxième année).

2.5 RECHERCHE CLINIQUE

Depuis 2010, je fais deux vacations hebdomadaires au CHU de Nantes comme ingénieur hospitalier en Méthodologie/Biostatistique au sein de la plateforme de Biométrie du CHU de Nantes. Cela m'a permis de travailler sur plusieurs études de recherche clinique, dans des domaines variés, et avec des méthodologies diverses. Ceci a donné lieu à plusieurs articles publiés dans le domaine de la bactériologie [5, 16] [74] [97] [44, 31], la génétique [15], de la maladie vasculaire [20, 17] ou de la médecine d'urgence [21] [46].

D'autres collaborations sont en cours par le biais d'études financées n'ayant pas encore donné lieu à des publications :

- En psychiatrie : Etude du profil des malades de Parkinson développant une addiction de type comportementale au cours du traitement dopaminergique : aspects psychopathologiques, neurologiques et pharmacocinétiques (projet PAR-KAAD) [ANR/programme SAMENTA 2012]
- En anesthésie : Evaluation médico-économique de la surveillance par oxymétrie cérébrale (INVOS) en peropératoire lors d'endartériectomies carotidiennes (projet EMOCAR) [STIC 2010]
- En chirurgie digestive : Evaluation médico-économique des nouveaux traitements chirurgicaux de la maladie hémorroïdaire : Essai randomisé multicentrique comparant la ligature élective des artères hémorroïdales sous contrôle Doppler avec mucopexie par voie trans-anale à l'hémorroïdopexie circulaire par agrafage selon la technique de Longo (projet LIGALONGO) [STIC 2010]
- En gynécologie, chirurgie ORL et chirurgie plastique : Evaluation médico-économique de la fermeture des plans sous-cutanés par l'utilisation d'une agrafeuse à agrafes résorbables (INSORB) (projet S2CARA) [STIC 2011]
- En psychiatrie : Suivi d'une cohorte française de joueurs (jeu de hasard et d'argent). Evolution de la pratique de jeu et recours aux soins (projet EVALJEU) [PHRC 2009]
- En chirurgie pédiatrique : Etude multicentrique nationale de la qualité de vie des patients atteints de malformations anorectales (projet MARQOL) [PHRC 2010]
- En psychiatrie : Etude clinique monocentrique, prospective, comparant deux groupes de joueurs problématiques ("jeux en ligne" versus "jeu hors ligne")

débutant des soins dans le service d'addictologie du CHU de Nantes (projet EVALJEU-en ligne) [PHRC 2011]

2.6 EPIDÉMIOLOGIE ET SANTÉ PUBLIQUE

Enfin, de 2001 à 2006, j'ai travaillé comme chargé de mission à l'Observatoire Régional de la Santé (ORS) du Centre (Orléans), ce qui m'a permis d'être en contact direct avec les structures régionales de Santé Publique et de travailler sur des études dans différents domaines comme la périnatalité [133, 134, 139, 142], le diabète [122, 123], l'alcoologie [137, 136] [82, 83], le suicide [132, 138], le cancer [126, 131], les personnes âgées [140] ou l'évaluation et l'organisation des soins [124, 125, 135, 141].

2.7 CONCLUSION

En faisant ce bilan, trois points m'ont interpellé pour la poursuite de mes travaux de recherche, car les différents travaux présentés mettent en évidence l'importance indéniable du modèle de Rasch et des études de simulations :

- Le modèle de Rasch présente-t-il un intérêt pratique indéniable, notamment est-il vraiment plus performant que l'approche basée sur un score (CTT) ?
- Pourquoi continuer sur le modèle de Rasch, est-ce que d'autres modèles de l'IRT ne pourraient pas présenter un plus grand intérêt ?
- Enfin, les travaux basés sur les études de simulations sont-ils les plus performants ? D'autres méthodologie sont-elles envisageables ?

2.7.1 Le modèle de Rasch présente-t-il un intérêt ?

Il me semble important face à ce bilan de s'interroger sur l'utilité du modèle de Rasch en Recherche Clinique et en Epidémiologie, face à l'alternative plus simple qu'est la CTT. Au regard des résultats présentés précédemment, le modèle de Rasch peut être vu comme un outil adapté à des comparaisons de groupes, mais peu adapté à une mesure individuelle : les analyses en deux temps basées sur le modèle de Rasch, c'est-à-dire en estimant dans un premier temps les valeurs individuelles du trait latent, puis en faisant des analyses sur ces valeurs individuelles, montrent des résultats décevants. En effet, ils induisant à la fois du biais et une perte de puissance comparées à l'approche CTT ou à l'introduction de covariables de groupes directement dans le modèle de Rasch (modèles de régression latente). Nous avons ainsi testé ces approches en deux temps dans une vision transversale [30] ou longitudinale [1], avec la même conclusion. C'est pourquoi dans des travaux plus récents que nous avons menés, ces approches ont été complètement abandonnées. C'est pourtant la seule approche envisageable dans un modèle de Rasch à effets fixes, où il n'est pas possible d'intégrer de paramètres de groupes dans le modèle et où, hormis la phase de validation d'un questionnaire, aucune analyse n'est possible sans passer par les estimations individuelles. Il est ici intéressant de noter que ces approches sont largement utilisées dans la littérature et dans les logiciels spécifiques liés au modèle de Rasch (ce type d'approche est par exemple à la base de toutes les analyses présentées dans le logiciel RUMM), et ce message risque donc d'être difficile à faire passer car il peut remettre en cause un certain nombre de résultats dans la littérature.

Une explication à cette mauvaise qualité de l'approche en deux temps tient à ce qu'elle entraîne trois types d'erreurs différentes sur les estimations individuelles du trait latent :

- D'une part, l'erreur liée à la fiabilité intrinsèque du questionnaire, qui peut se comprendre par le biais de la théorie de l'Univers d'Items (un concept latent peut être mesuré par une infinité d'items, et un questionnaire représente donc un échantillon de ces items, à partir de là, la fiabilité de la mesure que l'on obtient est liée au nombre d'items présents dans le questionnaire, et ne peut pas être parfaite, à moins d'imaginer un questionnaire contenant une infinité d'items).
- D'autre part, l'erreur liée au modèle de mesure (dans notre cas le modèle de Rasch) induite par l'erreur d'approximation sur les paramètres du modèle.
- Enfin l'erreur liée au modèle d'analyse (une ANOVA par exemple pour expliquer les valeurs individuelles par un effet groupe).

L'erreur globale est donc induite par l'addition de ces trois types d'erreurs. La moindre qualité par rapport à une approche basée sur un modèle de régression latente peut être due au fait que dans ce dernier modèle, il n'y a pas d'erreur lié au modèle d'analyse (l'analyse des effets groupes étant réalisée directement à partir des paramètres du modèle de mesure). L'approche basée sur un score (CTT) n'induit pas, quant-à elle, d'erreur liée au modèle de mesure, étant donné que la mesure, construite comme la simple somme des réponses aux items, n'induit pas de transformation par le biais d'un modèle.

En terme d'utilité, il semble que le modèle de Rasch présente un intérêt modéré lorsque les données manquantes sont absentes ou rares, mais qu'il présente des performances très intéressantes en présence de données manquantes. En effet, dans ce dernier cas, l'approche CTT présente de réelles limites, en permettant soit d'obtenir des estimations non biaisées des paramètres, mais une piètre puissance, lorsque l'analyse est réalisée en cas complets (en ne considérant que les individus n'ayant aucune données manquantes) ou bien en maintenant la puissance mais au prix d'un biais important sur les paramètres, lorsque l'on impute les données manquantes. L'approche basée sur un modèle de Rasch, dans lequel on ajoute des effets groupes semble alors l'approche la plus intéressante puisque qu'elle permet d'obtenir à la fois des paramètres non biaisés et d'être la méthode d'analyse la plus puissante pour détecter un effet groupe ou un effet temps. De plus, ces résultats restent valables quelle que soit la nature des données manquantes (informative ou non), et il semble donc peu utile d'utiliser des approches pour prendre en compte des données informatives, comme les modèles de sélection, ou le pattern mixture.

Ces résultats peuvent donc nous amener à penser que le modèle de Rasch pourrait être à la base de méthodes d'analyses puissantes pour la Recherche Clinique et l'Epidémiologie lorsque la variable d'intérêt est un PRO, où il s'agira d'étudier des effets sur des échantillons pour effectuer de l'inférence sur des populations, mais n'est pas une méthode à recommander pour une évaluation individuelle d'un PRO où un score sera plus efficace et moins compliquée. Néanmoins cela suppose que des ponts entre les deux types d'analyses soient posés, afin de permettre d'extrapoler des résultats obtenus en Recherche Clinique par exemple au niveau individuel, et donc de changer de métrique pour passer d'une mesure par intervalles à une mesure ordinale. Les travaux actuels que nous menons avec Alexandra Rouquette sur la définition d'une différence

minimale cliniquement pertinente (voir section 3.1.2) en se basant sur une variable latente vont dans ce sens, puisque pour pouvoir facilement interpréter les résultats sur le trait latent, nous faisons le parallèle avec le score. La propriété d'exhaustivité du score sur le trait latent du modèle de Rasch est aussi une facilité offerte par les modèles de la famille de Rasch d'effectuer ce lien, en créant une liaison bijective entre les deux métriques.

Ces résultats semblent justifier en soi de continuer à travailler sur les modèles de la famille de Rasch comme alternative intéressante à la CTT en Recherche Clinique et Epidémiologie.

2.7.2 Et les autres modèles de l'IRT

Durant ces dernières années, l'attention de l'EA 4275 s'est portée principalement sur les modèles de la famille de Rasch (modèles vérifiant les propriétés d'exhaustivité du score sur le trait latent et d'objectivité spécifique, famille incluant le modèle de Rasch et l'OPLM (One Parameter Logistic Model) pour les items dichotomiques, le Rating Scale Model et le Partial Credit Model pour les items polytomiques), qui sont des modèles particuliers de l'IRT. Néanmoins, un important nombre d'autres modèles existent en IRT, dont les plus connus sont le 1PLM (modèle logistique à 1 paramètre qui est en fait le même modèle que le modèle de Rasch, vu comme un cas particulier de modèle plus complexe, et sans perspectives sur ses propriétés spécifiques), le modèle de Birnbaum (2PLM modèle logistique à 2 paramètres) ou le 3PLM (modèle logistique à 3 paramètres), le Graded Response Model (GRM - modèle à réponses gradées), le Generalized Partial Credit Model (GPCM), ou les modèles à liens probit (Normal Ogive Models). Ces autres modèles restent néanmoins relativement peu utilisés en Recherche Clinique ou en Epidémiologie (ils le sont plus en sciences de l'éducation), bien que le GRM ou le GPCM soit parfois utilisés comme modèle sous-jacent à la construction de questionnaires basés sur du CAT (Computerized Adaptive Testing) pour l'analyse de PRO comme pour le projet PROMIS (Patient Reported Outcomes Measurement Information System).

On peut reprocher à nos travaux le fait de se restreindre à un modèle particulier de l'IRT, qui de plus s'adapte souvent mal à des données issues de questionnaires qui n'ont pas été construits selon ce modèle. En effet, l'utilisation de n'importe quel modèle de l'IRT permet d'obtenir une mesure d'intervalle pour le concept étudié, et les modèles plus complexes de l'IRT permettent généralement d'avoir de meilleures adéquations aux données que le modèle de Rasch. Néanmoins, le modèle de Rasch offre deux hypothèses qui nous semblent importantes et qui lui donnent sa suprématie sur les autres modèles. D'une part, l'exhaustivité du score sur le trait latent permet de justifier en soi que le score simple, généralement utilisé, est une bonne mesure ordinales du concept latent. Ceci permet donc d'éviter le fait qu'il faudrait interpréter des mêmes scores différemment en fonction des items qui seraient positifs, lors d'une utilisation pratique d'une échelle basée sur l'usage des scores. D'autre part, l'objectivité spécifique a montré ses réels avantages lors de nos différents travaux. Cette propriété induit que l'estimation du trait latent est indépendante des items répondus par l'individu, aussi, quelle que soit la raison pour laquelle un individu n'a pas répondu à certains items (inattention, fatigue, peur de répondre, choix personnel de ne pas répondre, CAT (cas où l'individu ne répond qu'aux items qui semblent les mieux adaptés à son état)...), les estimations produites restent non biaisées. C'est un résultat que nous avons retrouvé

plusieurs fois, dans différents cadres, et qui permet d'expliquer le fait que dans les analyses basées sur le modèle de Rasch, le niveau d'informativité des données manquantes n'influence pas les estimations des paramètres. Ceci permet une gestion très aisée des données manquantes, puisqu'il suffit de ne pas imputer les données manquantes et de travailler seulement sur les données disponibles (available case) pour obtenir des résultats satisfaisants. Cette propriété d'objectivité spécifique est une des spécificités des modèles de la famille de Rasch, et les autres modèles de l'IRT ne vérifient pas cette propriété. A ce titre, il est très étonnant que des projets en CAT comme PROMIS ne soient pas basés sur un modèle de la famille de Rasch. Ce thème reste d'ailleurs un grand sujet de discorde dans le milieu de l'IRT, entre les chercheurs défendant absolument le modèle de Rasch en mettant en avant les propriétés spécifiques des modèles de la famille de Rasch, quitte à devoir modifier le questionnaire pour que les données aient une bonne adéquation à ce type de modèle, et ceux pour qui l'essentiel reste l'idée d'estimer un concept latent par une mesure d'intervalle, qui vont privilégier des modèles plus souples, mais plus complexes et ne vont pas s'intéresser aux propriétés spécifiques du modèle de Rasch.

Dans le cas de nos travaux, il pourrait néanmoins être intéressant d'étendre les différentes études de simulations que nous avons réalisées à des modèles plus souples comme le 2PLM, afin de voir l'impact de l'informativité des données manquantes sur les analyses issues de ce modèle. Ceci nécessite néanmoins une réflexion importante sur la manière de simuler les données, de manière à créer certes des données différentes de celles obtenues avec le modèle de Rasch afin de justifier l'usage d'autres modèles, mais pas trop différentes non plus, de manière à pouvoir comparer les résultats à ceux obtenus avec le modèle de Rasch.

2.7.3 Les études de simulations de Monte Carlo

Un dernier point qui peut paraître surprenant dans ce bilan est le recours fréquent (voir quasi exclusif pour les travaux d'ordre méthodologique) aux études de simulations de Monte Carlo. Cette méthode offre en effet plusieurs avantages :

- la maîtrise parfaite des données produites (hypothèses sous jacentes, valeurs des paramètres...)
- la facilité à produire des résultats facilement interprétables (risque alpha, puissance, biais...)
- la possibilité d'étudier une grande variété de cas (que ce soit des cas réalistes, c'est-à-dire proche de données réelles, ou plus expérimentales en poussant les paramètres sur des valeurs extrêmes voire abérantes, de façon à approcher le comportement des méthodes testées dans des conditions extrêmes)
- la possibilité de réaliser un grand nombre de scenarii permettant une certaine généralisation des résultats
- la possibilité de réaliser des études de validation (en analysant les données simulées par le même modèle que celui ayant servi à la simulation) ou de robustesse (en simulant dans des conditions différentes que celles requises par le modèle d'analyse)

Néanmoins, ces méthodes de simulations ont un défaut qui est la difficulté à simuler correctement des données réalistes ressemblant le plus possibles à des données réelles, ce qui nécessite pour y arriver une grande réflexion sur la manière de simuler les données et une maîtrise de données réelles afin de fixer des valeurs réalistes pour les différents paramètres.

On pourrait s'interroger néanmoins sur le fait de ne pas utiliser d'alternatives pour les travaux méthodologiques, comme l'usage de données réelles, par exemple, en partant de données réelles pour créer différents scénarii, à l'aide de simulations ciblées sur certains paramètres ou variables (création d'une variable, création de données manquantes...). Cette méthode pose le problème du manque de maîtrise de la structure des données utilisées (quel modèle sous jacent aux données est réellement pertinent, les hypothèses faites sont-elles réalistes ou vérifiées, quel est le processus sous jacent aux données manquantes, quelles seraient les vraies valeurs des paramètres...). Cette méthode est de plus moins facilement généralisable car elle requiert une base fixe (les données réelles) à partir de laquelle seules certaines informations peuvent être manipulées. Cela ne semble donc pas une alternative aussi intéressante que les simulations de Monte Carlo où les données sont créées de toutes pièces.

Perspectives et Projets

Eu égard au bilan présenté précédemment, voici un résumé des perspectives à court et moyen terme de mes travaux de recherche.

3.1 DÉTERMINATION DU NOMBRE DE SUJETS NÉCESSAIRES POUR L'ANALYSE DE MESURES SUBJECTIVES

3.1.1 Le projet ANR PLANIPRO (2011-2013)

Le projet PLANIPRO vise à établir une méthode de détermination du nombre de sujets nécessaires en phase de planification d'un essai lorsque le critère principal est une mesure subjective mesurée par un modèle IRT, et que l'objectif vise à comparer des groupes de patients, ou à déterminer un effet temps dans une approche longitudinale.

Actuellement, une méthode a été proposée pour déterminer le nombre de sujets nécessaires quand le questionnaire ne comporte que des items dichotomiques, et avec deux groupes de patients. Une extension de ce travail est en cours pour les items polytomiques. En parallèle, ces méthodes sont en cours de validation dans des cas extrêmes, c'est-à-dire dans des conditions non optimales pour ce type d'analyse, par exemple en utilisant des items dont les difficultés seraient mal réparties, lorsque le questionnaire est mal adapté à la population visée (items globalement très faciles ou très difficiles), ou encore lorsque la population est très hétérogène (variance très grande du trait latent). Ces analyses sont aussi étendues pour vérifier la robustesse de la méthode lorsque les hypothèses du modèle de Rasch ne sont pas vérifiées, en particulier lorsque les items ne sont pas localement indépendants, en présence de fonctionnement différentiel des items (DIF) ou lorsque le trait latent n'est pas distribué normalement.

En outre, dans le cadre de ce projet, un ratio de correction du nombre de sujets nécessaires a été mis en évidence qui permettrait de déterminer la correction à effectuer sur le nombre de sujets nécessaires déterminés à partir de la formule classique (correcte pour des variables manifestes) afin de déterminer un nombre de sujets nécessaires adaptés à une variable latente. Ce ratio qui se calcule comme le rapport entre les deux nombres de sujets nécessaires (celui obtenu par la méthode issue de l'IRT et celui obtenu par la formule classique) pour obtenir une même puissance est constant quelle

que soit la valeur de la puissance choisie. Nous travaillons actuellement à découvrir les liens entre les différents paramètres (nombre d'items, paramètres de difficulté, variance du trait latent, effet groupe...) et la valeur de ce ratio, afin de proposer une manière de déterminer rapidement ce ratio (et donc de calculer le nombre de sujets nécessaires). Ceci nous semble une manière simple, efficace, rapide et facile à comprendre pour déterminer le nombre de sujets nécessaires dans ce contexte.

L'extension de ce travail à moyen et long terme ira dans différentes directions :

- L'extension de la méthode proposée au cadre longitudinal pour déterminer un effet temps, puis un effet groupe couplé à un effet temps
- L'impact d'erreurs de planification sur le calcul du nombre de sujets
- La correction d'un nombre de sujets nécessaires en cours d'étude (afin de tenir compte d'éventuelles erreurs de planification)
- La création de logiciels adaptés pour la détermination des nombres de sujets nécessaires dans ces différents contextes (logiciels indépendants, ou intégrés dans des logiciels statistiques généralistes)

Ce travail sera ensuite étendu aux analyses séquentielles, dans différents contextes (un ou deux groupes, cadre transversal ou longitudinal...).

3.1.2 L'interprétation de la différence sur le trait latent entre deux groupes de patients, pour définir la Différence Minimale Cliniquement Pertinente

Dans le cadre du projet PLANIPRO, le comité scientifique s'est posé la question du sens à donner à une différence sur le trait latent entre deux groupes ou deux temps, et donc de la manière de définir une différence minimale cliniquement pertinente. Ce sujet nécessite d'être creusé, et une première approche basée sur l'analyse de jeux réels a été décidée. Plusieurs jeux de données ont été ciblés :

- Un jeu de données provenant d'une cohorte de patients atteints de Broncho-pneumopathie Chronique Obstructive (BPCO) suivis au CHU de Nantes, et auxquels a été proposé dans le cadre de leur traitement un programme de réhabilitation fonctionnelle basé sur l'exercice physique. Ces patients sont vus tous les 6 mois la première année puis tous les ans. A chaque visite, un examen clinique est réalisé et il est complété par le remplissage du questionnaire de qualité de vie Saint Georges spécifique à cette pathologie. Une différence de 4 points sur le score Saint Georges est reconnue comme la différence minimale cliniquement pertinente. Une première étude réalisée sur cette cohorte visant de déterminer les facteurs pronostiques d'une amélioration de la qualité de vie du patient a montré que le niveau initial du score était un facteur important, et qu'il était plus facile d'améliorer la qualité de vie chez les patients ayant une faible qualité de vie initiale, que pour ceux ayant une qualité de vie moyenne ou correcte initialement. Cela a laissé présager que le score de qualité de vie de Saint Georges n'était pas une bonne mesure d'intervalle, et donc, il semble intéressant de voir si les items de ce questionnaire ne peuvent pas être analysés par un modèle de Rasch, afin de pouvoir travailler sur une mesure d'intervalle. Ainsi, il pourrait être envisagé de rechercher une différence de trait latent qui corresponde à une différence moyenne de 4 points du score, et de regarder ce que cette nouvelle différence peut représenter en fonction du niveau de score (bas, moyen, élevé). D'autres pistes sont aussi envisagées, par exemple l'utilisation de score clinique sur l'essoufflement (échelle de Sadoul cotée de 0 à 5) afin de voir le changement

- engendré sur le trait latent d'une modification de 1 point sur cette échelle (ce changement étant considéré comme une différence pertinente). Ce travail sur cette cohorte souffre néanmoins de plusieurs problèmes : d'une part, le nombre de patients suivi est faible (n=78) ; d'autre part, il s'agit de travailler sur une échelle de qualité de vie, et contrairement à ce qui est préconisé dans ce domaine, nous ne pouvons pas déceler une différence cliniquement pertinente évaluée par le patient, car cette information n'est pas disponible ; enfin, les données n'ont pas été relevées de façon optimale et la saisie souffre de défauts, en particulier, certaines données manquantes ont été à tort codées 0, comme la meilleure modalité de chaque item, avec une impossibilité *a posteriori* de différencier les données manquantes des réponses codées 0.
- Un jeu de données provenant d'une cohorte de joueurs de jeu de hasard et d'argent, à risque ou non, et suivis ou non (pour les joueurs à risque). Dans cette cohorte, un ensemble vaste de questionnaires a été rempli, incluant des questionnaires sur les pathologies psychiatriques, les tempéraments, la dépendance au jeu, les pratiques de jeu, les antécédents somatiques, et les distorsions cognitives liées au jeu. L'idée serait de travailler sur le questionnaire GABS (Gambling Attitudes and Beliefs Scale) qui mesure les distorsions cognitives liées au jeu, et pour lequel nous avons proposé une structure en 5 dimensions, et de chercher à définir sur chacune de ces dimensions les différences cliniquement pertinentes. L'avantage ici est de pouvoir se baser de façon plus satisfaisante sur une approche basée sur la comparaison des scores entre des groupes définis en fonction des variables cliniques. En effet, les joueurs ne pouvant pas eux-mêmes définir une amélioration (ou détérioration) importante de leurs distorsions cognitives, par conséquent, cette définition ne pourra être réalisée que par les médecins ou à l'aide de données indépendantes (contrairement aux patients atteints de BPCO qui sont capables d'évaluer eux-mêmes leur qualité de vie, et donc pour lesquels, il est moins satisfaisant de définir une différence cliniquement pertinente sur la base des experts ou de données indépendantes). La cohorte comporte plusieurs centaines de joueurs, néanmoins, nous n'avons actuellement à disposition que les variables issues du temps de recrutement, l'inclusion ayant eu lieu jusqu'en septembre 2011 (néanmoins, environ 80 patients ont eu une première visite à 1 an).
 - Un jeu de données sur les étudiants de PAES qui ont répondu au SF36 à deux temps d'étude (début d'année universitaire et après les résultats de la première partie du concours). L'avantage est ici d'avoir une base de données importante (625 étudiants) ayant répondu aux deux temps. La détection de la différence minimale cliniquement pertinente peut être faite indépendamment du SF36, en s'appuyant sur une évaluation sur une échelle de Likert mesurant la qualité de vie globale sur 7 niveaux. En outre, nous savons que la qualité de vie émotionnelle de ces étudiants s'est dégradée entre les deux passations du questionnaires, néanmoins, nous avons observé peu de scores extrêmes (très haut ou très bas), alors que nous savons que c'est sur ces niveaux extrêmes que l'IRT se distingue le plus de la CTT.
 - Enfin, un jeu de données provenant de l'étude SATISQOL, une étude multicentrique (CHU de Besançon, Marseille et Nancy) de 1520 patients hospitalisés en milieu médical et chirurgical pour pathologie chronique dans les domaines cardio-vasculaires, locomoteur, nephro-urologique, digestif, pneumologique, endocrinologique et bénéficiant d'un geste interventionnel pendant leur séjour. Ces

patients ont répondu au SF36 à trois temps de mesure (inclusion, 6 mois, 12 mois) et la question 2 du SF36 peut être utilisée pour détecter un changement (Par rapport à il y a 6 mois, comment trouvez-vous votre état de santé en ce moment ?)

Ces travaux sont envisagés dans le cadre de la thèse d'exercice de médecine d'Alexandra Rouquette, dont la soutenance devrait avoir lieu en 2013. Dans le cadre de cette thèse, une méthodologie sera proposée pour déterminer une différence minimale cliniquement pertinente en se basant sur le trait latent à l'aide des modèles de la famille de Rasch.

3.1.3 Le nombre de sujets nécessaires pour la validation d'échelles

Si le projet PLANIPRO aborde la problématique du nombre de sujets nécessaires pour comparer des groupes de patients ou pour déterminer une évolution dans le temps, il n'aborde pas cette problématique dans le cadre de la validation de questionnaires, que ce soit dans le cadre d'une validation classique d'un score, ou de la validation par un modèle IRT. Pourtant, il s'agit d'une problématique en soi, qui ne semble pas avoir trouvé de solution jusqu'à présent. On peut en effet trouver des recommandations, qui, selon les auteurs, peuvent donner des résultats sensiblement différents (du simple au quadruple) voire contradictoires (à nombre de dimensions constant, certains auteurs recommandent par exemple une augmentation du nombre de sujets nécessaires lorsque le nombre d'items augmente, alors que d'autres préconisent le contraire). Ce sujet reste donc d'actualité et sera notamment traité dans le cadre de la thèse d'Emmanuelle Anthoine, tout d'abord par le biais d'une revue systématique de la littérature, puis par le biais de simulations en se focalisant dans cette dernière partie sur la validation de structure à l'aide d'une analyse factorielle confirmatoire, et en se focalisant sur les indices d'adéquation les plus utilisés dans cette analyse. Cette thèse est prévue sur les 3 prochaines années (elle a commencé en octobre dernier), et actuellement la revue de la littérature est en cours.

En ce qui concerne la détermination du nombre de sujets nécessaires dans le cadre de la validation d'un questionnaire par un modèle IRT, je souhaiterais arriver à développer cette thématique dans les prochaines années, par exemple dans le cadre d'un doctorat. Il pourra en particulier s'agir d'étudier la puissance des tests d'adéquation, de manière à avoir un nombre de sujets suffisant pour réaliser ce type de tests dans des conditions satisfaisantes. Un projet est actuellement en cours de rédaction en ce sens pour obtenir un financement via une allocation de recherche, ou par le biais de la fondation de l'Université de Nantes.

3.2 VALIDATION D'ÉCHELLES

3.2.1 Validation de la version française de l'échelle Impact of Cancer v2

L'échelle "Impact of Cancer" mesure l'impact quotidien de la maladie pour les patients atteints d'un cancer. Ce questionnaire mesure 11 dimensions (47 items) dans trois domaines : les impacts positifs (épanouissement personnel, hygiène de vie, modifications existentielles, autoévaluation positive), les impacts négatifs (apparence, chan-

gements corporels, inférence dans la vie, peur de la récurrence), et l'impact sur l'emploi et les relations (emploi, relations hors conjoints, relation avec le conjoint). Ce questionnaire en langue anglaise a été traduit en français par l'EORTC mais sa traduction n'a pas été validée. L'institut Gustave Roussy, dans le cadre du protocole CANTO (Etude des toxicités chroniques des traitements anticancéreux chez les patientes porteuses de cancer du sein localisé - Coordinateur principal : Dr Fabrice André) prévoit la création d'une cohorte de 20 000 patientes atteintes d'un cancer du sein, et souhaite poser à ces patientes le questionnaire Impact of Cancer (le recrutement de la cohorte débutant en 2013). Ainsi, le Dr Sarah Dauchy, coordinatrice de l'axe "Qualité de vie, psychologie" de cette cohorte nous a demandé de réaliser l'étude de validation de ce questionnaire. Il est prévu d'inclure 300 femmes en 2012 dans le cadre de cette étude, le recrutement étant actuellement en cours. La validation de cette échelle sera effectuée par Myriam Blanchin et moi-même dans les prochains mois.

3.2.2 Validation de la version française de l'Index de la Fonction Sexuelle Féminine (IFSF)

L'échelle "Female Sexual Function Index - FSFI" a été développée en 2000 par Rosen et al. Cette échelle mesure six dimensions (Désir, Excitation, lubrification, Fonction Orgasmique, Satisfaction, Douleur) à l'aide de 19 items. Cette échelle a été traduite et validée par une équipe québécoise auprès des femmes de plus de 65 ans. Dans le cadre de sa thèse d'exercice sur l'impact de la dermatose vulvaire, Sophie Wylomanski (Interne en dermatologie - CHU de Nantes) a souhaité utiliser cette échelle. Néanmoins, elle souhaitait au préalable la valider auprès d'une population française de tous âges. Une étude de validation est donc en cours auprès des femmes consultant au service de Gynécologie du CHU de Nantes. Il est prévu le recrutement de 500 femmes sur le mois d'avril 2012 afin de valider cette échelle. La validation sera réalisée au sein de l'EA4275 lors d'un stage d'Internat par Antoine Vanier prévu de mai à octobre 2012 sous ma direction.

3.2.3 Validation du questionnaire de qualité de vie spécifiques aux maladies neuromusculaires

Depuis 2007, je travaille en collaboration avec l'EA3797 du CHU de Reims sur l'élaboration d'un questionnaire de qualité de vie adaptés aux patients atteints de maladies neuromusculaires (cf section 2.2.2). Une version longue du questionnaire est actuellement soumise à environ 120 patients, dans le but de sélectionner les items importants dans cette base de données et de proposer un questionnaire raisonnable d'environ 40 questions. Cette analyse devrait avoir lieu début 2012 et sera complétée par une validation du questionnaire finale fin 2012.

3.3 COMPARAISON DE L'ANALYSE DES MESURES SUBJECTIVES ENTRE LA CTT ET L'IRT

3.3.1 Etudes transversales

La comparaison des approches CTT et IRT dans un cadre longitudinal dans l'idée de comparer des groupes de patients va être poursuivie dans les prochaines années, notamment par le biais de la thèse de Jean-François Hamel. Quatre axes importants

pourront être creusés : d'une part, la présence de Fonctionnement Différentiel des Items (DIF) entre des groupes de patients (différents ou non des groupes que l'on cherche à comparer), et la prise en compte ou non de ce DIF ; d'autre part la quantification de l'apport d'une ou plusieurs covariables pour expliquer un trait latent ; ensuite l'extension à des items polytomiques ; enfin l'extension de ce travail à d'autres modèles de l'IRT, en particulier à un modèle de Birnbaum (2PLM).

En ce qui concerne le DIF, celui-ci est certes courant mais est rarement pris en compte en CTT, et n'est pas toujours pris en compte en IRT. D'un autre côté, sa prise en compte en IRT consistant à utiliser des valeurs de paramètres liés aux items différents dans un groupe ou dans l'autre, pourrait avoir peu d'importance en ce qui concerne les tests de comparaison de groupes. En effet, au regard de résultats précédents, publiés notamment dans *BMC Medical Research Methodology* [22], le choix de valeurs biaisées pour les paramètres associés aux items avaient peu d'impact sur ce type de test, jusqu'à un certain niveau (assez important néanmoins) de biais. Dans le cadre du DIF, le contexte est un peu différent, car il ne s'agit pas forcément d'un biais sur ces paramètres, mais de différences sur les valeurs entre des groupes de patients. En outre, ces valeurs de paramètres d'items sont alors différentes d'un groupe à l'autre (contrairement à l'article cité, où le biais créé était présent pour tous les individus). Néanmoins, on peut se questionner sur l'intérêt de la prise en compte ou non du DIF sur un test comparant des groupes de patients, à moins d'un DIF très important.

En ce qui concerne la quantification des effets groupes, il est en effet important, outre le fait de mettre en évidence des effets significatifs ou non, d'avoir une idée de l'importance de la prise en compte de ces effets pour expliquer le trait latent. On peut en ce sens faire un parallèle entre les modèles IRT incluant des covariables pour expliquer le trait latent et les modèles linéaires classiques, puisque l'idée en ajoutant des covariables consiste à créer un modèle linéaire, au sein même du modèle de mesure, afin d'expliquer le trait latent, par les différentes covariables. Par analogie, il peut donc être intéressant de quantifier l'impact en terme de variance expliquée par les covariables. En effet, il est aisé de s'apercevoir dans cette approche que le trait latent peut être décomposé dans ce type de modèle en une partie expliquée (par les covariables analysées) et une partie inexpliquée (variable aléatoire représentant une sorte de résidu dans le modèle expliquant le trait latent). La décroissance de la variance de ces résidus permettrait donc de quantifier l'impact des covariables, et en comparant des modèles emboîtés (avec toutes les covariables en les retirant une à une à chaque fois (analyse de type III) ou bien en les ajoutant une à une (analyse de type I)), il sera possible de définir l'apport de chaque covariable.

L'extension à des modèles polytomiques est maintenant indispensable et sera une priorité de ce thème. Des programmes sous Stata pour réaliser ce type d'analyse en transversal sont actuellement en cours de validation dans l'équipe. Enfin, comme présenté dans le bilan, il sera intéressant de confronter ces résultats à une étude de simulation basée sur le modèle de Birnbaum, afin de voir les spécificités des modèles de la famille de Rasch par rapport à d'autres modèles de l'IRT.

3.3.2 Etudes longitudinales

L'analyse des données subjectives longitudinales sera poursuivie, en particulier dans le cadre de la thèse d'Elodie De Bock-Dumas, qui travaille sur l'impact des données manquantes intermittentes sur les items (fait qu'à chaque temps, une partie des réponses aux items ne soit pas observée) sur les propriétés des tests basés sur une analyse classique (CTT) ou sur un modèle de Rasch (IRT). Dans le cadre de ce travail, une des principales difficultés réside dans le temps nécessaire pour l'exécution des programmes, déjà sensible en présence de 3 temps de mesure, cette difficulté devient gênante pour la réalisation d'études de simulations portant sur 4 temps ou plus. L'idée sera donc dans un premier temps de travailler sur ce point afin d'optimiser les temps de calcul. Deux pistes sont envisagées :

- d'une part, une estimation des paramètres à l'aide de programmes de plus bas niveau que les langages de programmations des logiciels statistiques comme SAS, R ou Stata (que l'on pourraient classer comme langages de 4e génération ou de très haut niveau), de manière à améliorer l'optimisation des programmes. En ce sens, un travail sur FORTRAN (langage de haut niveau) est actuellement en cours d'évaluation en collaboration avec Etienne Dantan (EA 4275) pour réaliser l'algorithme d'estimation des paramètres des modèles IRT longitudinaux par un code FORTRAN dans nos études de simulation.
- d'autre part, un travail sur une manière de trouver rapidement des valeurs initiales des paramètres correctes, de manière à réduire le nombre d'itérations dans les algorithmes d'estimation.

Sur un plan plus pratique, suite à la première étude de simulations décrite précédemment, une seconde étude est prévue de manière à évaluer l'impact des données manquantes sur les tests associés à des effets groupes dans un cadre longitudinal. L'extension à des items polytomiques est dans ce contexte aussi un travail important à réaliser en priorité, qui devra être entrepris en parallèle du travail sur l'optimisation des programmes. Enfin, à moyen terme, le response-shift, c'est-à-dire la modification de la perception du concept mesuré au fil du temps (qui se traduit par une sorte de DIF temps) sera étudié. ce dernier point permettra de faire un lien avec les collègues de Sciences Humaines et Sociales de l'EA 4275 qui travaillent sur les processus d'adaptation des patients à la maladie (dans le contexte du cancer), qui pourrait s'apparenter, sur certains aspects à du response-shift.

3.4 AUTRES PERSPECTIVES À MOYEN TERME

3.4.1 Création d'un outil de caractérisation des pharmacodépendances médicamenteuses

Ce projet porté par Caroline Victorri-Vigneau (CEIP - CHU de Nantes) auquel je suis associé vise à déterminer des indices statistiques substance par substance afin de caractériser le potentiel de pharmacodépendance de chaque substance, à partir des notifications substances reçues par le CEIP, et en particulier à partir des réponses au questionnaire de mesure de la pharmacodépendance en cours de validation. L'outil reposera en particulier sur le graphique représentant en abscisse le nombre cumulés d'items positifs (moins de 1, moins de 2... moins de 8) et en ordonnée la proportion de notifications concernées. Sur cette courbe, une augmentation tardive montre une forte

proportion de réponses positives aux huit items composant le score, signifiant un fort niveau de dépendance de la substance, alors qu'une montée plus précoce représentera des substances avec un potentiel de dépendance moins important. La représentation 3D (avec le pourcentage cumulé de notifications concernés en fonction des deux sous-scores de l'échelle de pharmacodépendance) sera aussi étudié pour voir son intérêt par rapport à cette première courbe. Des indicateurs numériques seront créés en s'appuyant sur cette courbe (aire sous la courbe ou volume sous la surface, mesure de distance entre les courbes ou entre les surface,...) et interprétés pour pouvoir caractériser les potentiel de dépendance de chaque substance concernée par des notifications. Le projet a été soumis à l'appel d'offre de l'Affsaps en 2012.

3.4.2 Recherche Clinique

Je participe actuellement à plusieurs projets de recherche clinique qui sont actuellement en cours d'évaluation ou de conception et devraient commencer dans les prochains mois :

- En gérontologie : Etude multicentrique de soins courants, non invasive, en ouvert, prospective, visant à évaluer une nouvelle méthode de mesure de la rigidité artérielle
- En psychiatrie : modérateurs de jeux en ligne : quelle efficacité ? Etude expérimentale contrôlée et randomisée (projet MOD&JEU)
- En médecine d'urgence : Etude multicentrique, randomisée, comparant deux stratégies pour diminuer la consommation d'antibiotiques dans les pneumonies aiguës communautaires : une stratégie basée sur la réévaluation clinique et une stratégie guidée par la PCT (financement PHRC Interrégional 2012)
- En chirurgie digestive : Evaluation médico-économique d'un sphincter anal magnétique chez les patients atteints d'incontinence fécale sévère (projet MOS)
- En chirurgie endocrinienne : Etude prospective, multicentrique et observationnelle évaluant la qualité de vie après thyroïdectomie par cervicotomie
- En chirurgie endocrinienne : Symptômes non spécifiques et qualité de vie dans l'hyperparathyroïdie primaire modérée. Etude prospective, multicentrique, non randomisée, avant et après chirurgie.
- En médecine vasculaire : Maladie Veineuse thromboembolique chez le médullo-lésé après 3 mois.

3.4.3 Enseignement : le parcours M1 Bioinformatique-Biostatistique du Master Biologie Santé

Comme autre perspective importante dans les prochains mois, je vais prendre la co-responsabilité du parcours de niveau M1 en Bioinformatique et Biostatistique dans le cadre du Master Biologie-Santé de l'Université de Nantes, avec le Dr Christine Sinoquet (UFR des sciences - département d'informatique). Ce nouveau parcours qui ouvrira à la rentrée 2012 développe le parcours de Bioinformatique pré-existant, et sera la principale porte d'entrée sur l'Université de Nantes pour les étudiants scientifiques vers les spécialité de M2 en "Bioinformatique" et en "Modélisation en Pharmacologie Clinique et Epidémiologie". Ce parcours aura la spécificité de donner un enseignement équilibré entre les deux domaines que sont la Bioinformatique et la Biostatistique, par rapport à d'autres parcours existant au niveau national. Ceci nécessitera de renforcer le potentiel nantais en spécialistes du traitement de données en grandes dimensions

qui sera à la base de l'enseignement dispensé dans ce parcours et de créer des liens avec le tissu industriel et de recherche local intéressé par une telle formation.

Chapitre 4

Cinq publications représentatives du travail de recherche

4.1 TOWARDS POWER AND SAMPLE SIZE CALCULATIONS FOR THE COMPARISON OF TWO GROUPS OF PATIENTS WITH ITEM RESPONSE THEORY MODELS

J.-B. Hardouin, S. Amri, M.-L. Feddag, and V. Sébille. Towards power and sample size calculations for the comparison of two groups of patients with Item Response Theory models. *Statistics in Medicine* [IF=2.33], November 2011.

4.2 NONPARAMETRIC ITEM RESPONSE THEORY USING STATA

J.-B. Hardouin, A. Bonnaud-Antignac, and V. Sébille. Nonparametric Item Response Theory using Stata. *Stata Journal* [IF=2.00], 11(1) :30-51, 2011.

4.3 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 SIMULATION STUDY

V. Sébille, J.-B. Hardouin, T. Le Néel, G. Kubis, F. Boyer, F. Guillemin, and B. Falissard. 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 simulation study. *BMC Medical Research Methodology* [IF=2.15], 10 :24, 2010.

4.4 COMPARISON OF CTT AND RASCH-BASED APPROACHES FOR THE ANALYSIS OF LONGITUDINAL PATIENT REPORTED OUTCOMES

M. Blanchin, J.-B. Hardouin, T. Le Néel, G. Kubis, C. Blanchard, E. Mirallié, and V. Sébille. Comparison of CTT and Rasch-based approaches for the analysis of longitudinal patient reported outcomes. *Statistics in Medicine* [IF=2.33], 30(8) :825-838, April 2011.

4.5 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

J.-B. Hardouin, R. Conroy, and V. Sébille. 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 Medical Research Methodology* [IF=2.15], 11 :105, 2011.

Fiche de synthèse des encadrements et publications associées

5.1 ENCADREMENT DE THÈSE

- **Myriam Blanchin** (coencadrement à 50%) [soutenue le 10 juin 2011] *Développement de modèles de réponse aux items (modèles IRT) pour l'analyse de mesures subjectives en santé évaluées dans le cadre d'études longitudinales*
 - M. Blanchin, J.-B. Hardouin, T. Le Néel, G. Kubis, C. Blanchard, E. Mirallié, and V. Sébille. Comparison of CTT and Rasch-based approaches for the analysis of longitudinal Patient Reported Outcomes. *Statistics in Medicine* [IF=2.33], 30(8) :825-838, April 2011.
 - M. Blanchin, J.-B. Hardouin, T. Le Néel, G. Kubis, and V. Sébille. Analysis of longitudinal Patient-Reported Outcomes with informative and non-informative dropout : Comparison of CTT and Rasch-based methods. *International Journal of Applied Mathematics & Statistics (IJAMAS)* [IF inconnu], 24(SI-11A), 2011.
 - M. Blanchin, J.-B. Hardouin, T. Le Néel, G. Kubis, and V. Sébille. CTT and IRT-based approaches for group comparison of longitudinal patient reported outcomes : Impact of dropout mechanisms. *Soumis*, 2011.
- **Jean-François Hamel** (M2 Modélisation en Pharmacologie Clinique et Epidémiologie (Nantes) et doctorat : coencadrement à 60%) [doctorat débuté en octobre 2010] *Etude de différentes approches permettant la comparaison de groupes d'individus par l'analyse de mesures subjectives en santé en présence de données manquantes potentiellement informatives*
 - J.-F. Hamel, J.-B. Hardouin, T. Le Néel, G. Kubis, Serge Fanello, and V. Sébille. Study of different methods for comparing groups by analysis of subjective health measurements. *En révision dans PLoS ONE*, 2011.
 - J.-F. Hamel, V. Sébille, T. Le Néel, G. Kubis, and J.-B. Hardouin. Rasch models to detect a group effect for Patients Reported Outcomes with possibly informative missing data. *Soumis*, 2011.
- **Elodie De Bock-Dumas** (coencadrement à 60%) [doctorat débuté en janvier 2011] *Identification de stratégies d'analyse de variables latentes longitudinales en présence de données manquantes potentiellement informatives*
 - E. Dumas, J.-B. Hardouin, M. Blanchin, T. Le Néel, G. Kubis, and V. Sébille. Analysis of longitudinal Patient-Reported outcomes with inter-

mittent missing data (informative and non informative) : comparison of score or Rasch-based methods. *En cours d'écriture*, 2011.

5.2 ACCOMPAGNEMENT SCIENTIFIQUE DE DOCTORANT (SANS STATUT D'ENCADRANT)

- **Fanny Feuillet** [doctorat débuté en octobre 2011] *Apport des analyses en variables latentes en pharmaco-épidémiologie*
 - F. Feuillet, L. Bellanger, J.-B. Hardouin, C. Victorri-Vigneau, and V. Sébille. On comparison of clustering methods for pharmacoepidemiological data. *Soumis*, 2011.
- **Gaëlle Bouju** [doctorat débuté en octobre 2011] *Les jeux de hasard et d'argent : vers une approche spécifique des différentes formes de jeux*
 - G. Bouju, J.-B. Hardouin, C. Boutin, P. Gorwood, J.-D. Le Bourvellec, F. Feuillet, J.-L. Vénisse, and M. Bronnec. A shorter and multidimensional version of the gambling beliefs and attitudes survey (GABS-23). *Soumis*, 2011.
- **Emmanuelle Anthoine** [doctorat débuté en octobre 2011] *Détermination du Nombre de Sujets Nécessaires (NSN) dans le cadre de la validation classique d'un score type Patient Reported Outcomes (PRO)*.

5.3 ENCADREMENT DE MASTER 2 AVEC PRODUCTION SCIENTIFIQUE

- **Estelle Duclay** (coencadrement à 40%) [Master 2 Modélisation en Pharmacologie Clinique et Epidémiologie (Nantes) - 2011] *Influence de l'absentéisme des personnels soignants sur la satisfaction du patient hospitalisé*
 - E. Duclay, J.-B. Hardouin, V. Sébille, and L. Moret. Impact of staff absenteeism on patient satisfaction : an exploratory study in a french hospital. *En cours d'écriture*, 2012.
- **Pauline Clermidi** (coencadrement à 80%) [Master 2 Modélisation en Pharmacologie Clinique et Epidémiologie (Nantes) 2010] *Validation de la traduction française du questionnaire de qualité de vie spécifique HAQL auprès de patients de 6 à 11 ans atteints de malformations ano-rectales*
 - P. Clermidi, G. Podevin, C. Crétolle, Sarnacki S., and J.-B. Hardouin. The challenge of measuring quality of life in children with Hirschsprung's diseases or anorectal malformation. *En cours d'écriture*, 2011.
- **Fanny Feuillet** (coencadrement à 50%) [Master de Biostatistique (Bordeaux) 2009] et **Laura Wainstein** (coencadrement à 20%) [Master 2 Méthodologie et Statistique en Recherche Biomédicale (Paris 11) 2009] *Caractérisation pharmaco-épidémiologique de patients consommateurs de médicaments psychotropes : Apport des modèles à classes latentes*
 - L. Wainstein, C. Victorri-Vigneau, V. Sébille, J.-B. Hardouin, F. Feuillet, J. Pivette, A. Chaslerie, and P. Jolliet. Pharmacoepidemiological characterization of psychotropic drugs consumption using a latent class analysis. *International Clinical Psychopharmacology* [IF=2.76], 26(1) :54-62, January 2011.

- **Julie Léger** (Master de Biostatistique (Bordeaux) coencadrement à 50%) [2008] *Analyse longitudinale de données de qualité de vie de patientes atteintes du cancer du sein : influence du coping et des caractéristiques des patientes*
 - A. Bonnaud-Antignac, J.-B. Hardouin, J. Léger, A. Hartmann, F. Dravet, and V. Sébille. Quality of life and coping of women treated for breast cancer and their caregiver. What are the interactions? *accepté dans Journal of Clinical Psychology in Medical Settings* [IF=1.506], sous presse, 2012.

5.4 AUTRES ENCADREMENTS D'ÉTUDIANTS AVEC PRODUCTION SCIENTIFIQUE

- **Jean-Damien Le Bourvellec** (encadrement à 100%) [Licence Pro Statistique et Informatique pour la Santé (Bretagne Sud)- 2009] *Mesure des distorsions cognitives des joueurs pathologiques : validation en langue française des questionnaires GABS et GRCS*
 - M. Bronnec, G. Bouju, J.-L. Vénisse, J.-D. Le Bourvellec, and J.-B. Hardouin. Psychometrics validation of the french version of the Gambling Related Cognition Scale (GRCS). *Accepté pour publication dans Journal of Gambling Issues* [IF=1.303], 2012.
 - G. Bouju, J.-B. Hardouin, C. Boutin, P. Gorwood, J.-D. Le Bourvellec, F. Feuillet, J.-L. Vénisse, and M. Bronnec. A shorter and multidimensional version of the gambling beliefs and attitudes survey (GABS-23). *Soumis*, 2011.
- **Sarah Amri** (coencadrement à 50%) [ENS Cachan - Antenne de Bretagne (Bruz) - 2009] *Estimation d'un effet groupe dans le cadre du modèle de Rasch à effet aléatoire : approche numérique de l'erreur standard de l'estimation*
 - J.-B. Hardouin, S. Amri, M.-L. Feddag, and V. Sébille. Towards power and sample size calculations for the comparison of two groups of patients with Item Response Theory models. *Statistics in Medicine* [IF=2.33], November 2011.

5.5 AUTRES ENCADREMENTS PRÉVUS OU N'AYANT PAS ENCORE DONNÉ LIEU À UNE VALORISATION SCIENTIFIQUE

- **Mathilde Quere** (coencadrement à 100%) [DUT STID (Bretagne Sud) - 2011] *Analyse d'un questionnaire de mesure du stress et de la qualité de vie des étudiants de première année commune aux études de santé*
- **Alexandra Rouquette** (encadrement à 100%) [thèse d'exercice de médecine (Angers) 2011-2013] *Détermination d'une différence cliniquement pertinente sur le questionnaire GABS23 (Gambling Attitudes and Beliefs Scale - 23 items) auprès d'un échantillon de 609 joueurs*
- **Emilie Peron** (coencadrement à 40%) [Master 2 Modélisation en Pharmacologie Clinique et Epidémiologie (Nantes) - 2012] *Facteurs déterminants de la prescription de 6 IRSS pour des patients naïfs de tout traitement antidépresseur*
- **Gaëlle Romain** (coencadrement à 50%) [Licence Pro Statistique et Informatique en Santé (Bretagne Sud)- 2012] *Etude d'une cohorte de 623 joueurs de jeux de hasard et d'argent : analyses des différentes formes de jeux*

- **Antoine Vanier** (coencadrement à 50%) [Stage d'Internat de Santé Publique - 2012] *Validation de l'Index de la Fonction Sexuelle Féminine (IFSF) auprès de 500 femmes consultant en Gynécologie*