



Recherche de sous-groupes évolutifs et leur impact sur la survie : application à des patients atteints d'ataxie spinocérébelleuse (SCA)

Alhassane Diallo

► To cite this version:

Alhassane Diallo. Recherche de sous-groupes évolutifs et leur impact sur la survie : application à des patients atteints d'ataxie spinocérébelleuse (SCA). Santé publique et épidémiologie. Sorbonne Université, 2018. Français. NNT : 2018SORUS447 . tel-02924998

HAL Id: tel-02924998

<https://theses.hal.science/tel-02924998>

Submitted on 28 Aug 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE DE DOCTORAT DE
SORBONNE UNIVERSITE

Spécialité

Epidémiologie Clinique

ECOLE DOCTORALE PIERRE LOUIS DE SANTE PUBLIQUE A PARIS :
EPIDEMIOLOGIE ET SCIENCES DE L'INFORMATION BIOMEDICALE (ED 393)

Présentée par

M. Alhassane Diallo

Pour obtenir le grade de

DOCTEUR de SORBONNE UNIVERSITE

Sujet de la thèse :

Recherche de sous-groupes évolutifs et leur impact sur la survie : application à des patients atteints d'ataxie spinocérébelleuse (SCA)

Soutenance prévue le **24 Octobre 2018 à 14H**

Devant le jury composé de :

Mme. Sophie Tezenas du Montcel	MCU-PH	Directrice de thèse
Mme. Hélène Jacqmin-Gadda	DR	Rapporteuse
M. Cyril Goizet	PU-PH	Rapporteur
Mme. Laurence Meyer	PU-PH	Examinatrice
M. Jean-Christophe Corvol	PU-PH	Examineur
M. Raphael Porcher	MCU-PH	Examineur

*Je dédie ce manuscrit de thèse à
mon très cher Papa :
Diallo Ibrahima V*

Remerciements

En premier lieu, je remercie tout naturellement Dieu le Tout Puissant, Le très Miséricordieux pour tous les bienfaits qu'il m'a accordés.

A mes parents Diallo Ibrahima V (papa chéri) et Tanti Binta (ma nènè) : aucune formulation à ma connaissance ne peut exprimer la sensation de plénitude qui m'accompagne. Par la grâce de Dieu, je vous dois TOUT et j'espère que ce travail dissipera toutes vos peines. Je vous serai reconnaissant pour toujours. Que le bon Dieu vous garde encore aussi longtemps parmi nous et vous accorde une santé de fer.

Ma directrice de thèse, le Dr Sophie Tezenas du Montcel pour ta bienveillance, ta disponibilité, ta rigueur scientifique et ta gentillesse. J'espère de que notre collaboration continuera.

Mes rapporteurs, les Professeurs Hélène Jacqmin-Gadda et Cyril Goizet, j'ai été honoré que vous ayez accepté de critiquer mon travail et de faire parti de mon jury de thèse. Vos connaissances approfondies à la fois des modèles statistiques pour l'un et de la physiopathologie des maladies neurodégénératives en particulier des ataxies pour l'autre ont sans doute contribué à l'amélioration de ce manuscrit de thèse.

Je remercie également les examinateurs de mon jury : Pr Laurence Meyer, je suis honoré que vous fassiez parti de mon jury de thèse. Vous êtes la toute première à m'encadrer, merci infiniment pour votre rigueur scientifique, votre soutien et votre humanisme. Pr Jean-Christophe Corvol et Dr Raphael Porcher merci à vous de faire parti de mon jury.

A mon jumeau Diallo Alhousseiny, je n'oublierai jamais tous les bons moments que nous avons partagés ensemble. Ce travail est le notre. Je te remercie infiniment pour ton soutien moral et tes conseils.

A mes sœurs Dalanda, Rama, Mariama, Kadiatou, N'ta Bijou et Aicha Mady, à mes Tantes (Diaraye, Haby, Kadiatou, Nènen, N'Gayorè, Houssai Kane), à ma belle mère Hadja Kadiatou Soumah, à tonton Aliou, à mes oncles (les 2 Kaou Ibrahima, Thierno, Thome, Bappa Bora), à tous mes neveux et nièces, à ma grand-mère Nen Biro, à mes cousins (Ibro, Pétédji, Abdourahim, Dr Hassimiou, Amirou), à mes amis de toujours Modjos, kamissoko, Fréro Lille, Chouaiban, Yousseuf, Grand Cellou, Younous, Hamid, Djouldé et Hassimiou merci infiniment pour votre soutien.

Merci aux internes pour les bons moments passés ensemble et à toute l'équipe de BIOSPIM.

Enfin, merci à Marie mon amour. Merci pour ton soutien, ta confiance et ton amour.

Résumé

Résumé en Français

Recherche de sous-groupes évolutifs et leur impact sur la survie : application à des patients atteints d'ataxie spinocérébelleuse (SCA)

Dans les études de cohorte, le plus souvent les modèles utilisés supposent que la population d'étude suit un profil moyen d'évolution. Cependant dans de nombreux cas, comme pour les ataxies spinocérébelleuses (SCA), il n'est pas rare qu'une hétérogénéité soit suspectée. Cette hétérogénéité pourrait aussi être influencée par d'autres événements intercurrents par exemple l'évolution conjointe d'un second phénotype ou la survenue d'un événement telle que la sortie d'étude ou le décès. Dans la première partie de cette thèse, nous avons analysé l'évolution de l'IMC des patients SCA et recherché des profils d'évolution différente de l'IMC. Nous avons identifié 3 sous-groupes d'évolution de l'IMC : diminution (23% des patients), augmentation (18%) et stable (59%); et nous avons montré que les patients qui baissent leur IMC progressent plus rapidement. Dans la seconde partie, nous avons étudié la survie des patients SCA et développé un nomogramme pronostique. Nous avons montré qu'elle est différente selon le génotype. La survie est plus courte chez les SCA1, intermédiaire chez les SCA2 et SCA3, et plus longue chez les SCA6. Enfin, nous avons évalué l'impact à long-terme de la progression de l'ataxie sur la survie. Nous avons montré que la progression de l'ataxie est associée à une survie plus courte quel que soit le génotype. Uniquement chez les patients SCA1, nous avons identifié trois sous-groupes de patients homogènes en termes de progression de la maladie et de risque de décès.

Mots clés : [Ataxie Spinocérébelleuse, Modèles mixtes à classes latentes, Modèle de Cox, Nomogramme, Modèles conjoints, Prédictions dynamiques].

Résumé en Anglais

Research of sub-groups evolutions and their impact on survival : Application to Spinocerebellar Ataxia patient's

In cohort studies, most often the models used assume that the study population follows an average pattern of evolution. However, in many cases, such as for spinocerebellar ataxias (SCA), it is not uncommon for heterogeneity to be suspected. This heterogeneity could also be influenced by other intercurrent events for example joint evolution of a second phenotype or occurrence of an event such as dropout or death. In the first part of this thesis, we analyzed the evolution of the

BMI of SCA patients and looked for different evolution profiles of BMI. We identified 3 subgroups of BMI evolution : decrease (23% of patients), increase (18%) and stable (59%); and we have shown that patients who lower their BMI are faster disease progression. In the second part, we studied the survival of SCA patients and developed a prognostic nomogram. We have shown that it is different according on the genotype. Survival is shorter in SCA1, intermediate in SCA2 and SCA3, and longer in SCA6. Finally, we assessed the long-term impact of ataxia progression on survival. We have shown that progression of ataxia is associated with shorter survival regardless of genotype. Only in SCA1 patients, we identified three subgroups of homogeneous patients in terms of disease progression and risk of death.

Keywords : [Spinocerebellar ataxia, Latent Class Mixed models, Cox models, Nomogram, Joint models, Dynamic predictions]

Production scientifique dans le cadre de la thèse

Articles

1) **Diallo A**, Jacobi H et al. Body Mass Index decline is related to Spinocerebellar Ataxia disease progression. *Mov. Disord. Clin. Practice*. 2017 October ; 4(5) : 689-697.

2) **Diallo A**, Jacobi H et al. Survival in patients with spinocerebellar ataxia types 1, 2, 3 and 6 (EUROSCA) : a longitudinal cohort study. *Lancet Neurology*. 2018 April ; 17 : 327-34.

3) **Diallo A**, Jacobi H et al. Prediction survival with long-term disease progression in spinocerebellar ataxia types 1, 2, 3 and 6. En cours de soumission.

Présentation orale

La diminution au cours du temps de l'IMC est liée à la progression de l'ataxie spinocérébelleuse (SCA). EpiClin 11/24ème Journées des Statisticiens des Centres de Lutte Contre le Cancer, Saint-Etienne 2017.

Notations et Abréviations

Notations

- $\mathcal{L}$: Log-vraisemblance
- G : Nombre de classes latentes
- B : Matrice de variance-covariance des effets aléatoires
- b_i : Effets aléatoires
- Y : Marqueur longitudinal
- $h(\cdot)$: Fonction de risque instantanée
- $A(\cdot)$: Fonction de risque cumulée
- $S(\cdot)$: Fonction de survie
- T : Temps d'évènement

Abréviations

- *ACAD* : Ataxies Cérébelleuses Autosomiques Dominantes
- *ARN* : Acide Ribonucléique
- *CAG* : Répétitions d'une séquence d'acides aminés composés de : Cystéine Alanine Glycine
- *CRC – SCA* : Clinical Research Consortium for Spinocerebellar Ataxias
- *DRPLA* : dentatorubro-pallidoluysian atrophy
- *EQ – 5D* : EuroQol five Dimensions questionnaire

- *EUROSCA* : European study for Spinocerebellar Ataxias
- *IMC* : Indice de Masse Corporel
- *INAS* : Inventory of Non-Ataxia Signs
- *M* : Marginal
- *PHQ – 9* : Patient Health Questionnaire-9 score
- *SARA* : Scale for the Assessment and Rating of Ataxia
- *SCA* : Spinocerebellar ataxia
- *SCAFI* : SCA Functional Index
- *SS* : Spécifique aux Sujets
- *UHDRS* : Unified Huntington’s Disease Rating Scale

Table des matières

Remerciements	3
Résumé	4
Résumé en Français	4
Résumé en Anglais	4
Production scientifique dans le cadre de la thèse	6
Articles	6
Présentation orale	6
Notations et Abréviations	7
Notations	7
Abréviations	7
1 Introduction générale	11
1.1 Contexte	11
1.2 Problématique	11
1.3 Objectifs	12
2 Etat des connaissances	13
2.1 Généralités	13
2.1.1 Définition	13
2.1.2 Épidémiologie	13
2.1.3 Mécanismes physiopathologiques	14
2.1.4 Signes cliniques	15
2.1.5 Signes radiologiques	16
2.1.6 Diagnostic	16
2.2 Cohorte EUROSCA	16
2.2.1 Mesures de la sévérité de l'ataxie	17
2.2.2 Etude de l'histoire naturelle de l'ataxie	19
3 Méthodologie utilisée au cours de la thèse	22
3.1 Méthodes pour les données longitudinales	22
3.1.1 Modèles linéaires mixtes	22
3.1.2 Modèles mixtes à classes latentes	23
3.2 Modèles conjoints	26
3.2.1 Modèle conjoint à effets aléatoires partagés	26

3.2.2	Modèle conjoint à classes latentes	28
4	Analyse de trajectoire : Impact de la perte de poids sur la progression de l'ataxie	32
4.1	Résumé	32
4.1.1	Rationnel	32
4.1.2	Méthodes	32
4.1.3	Résultat	35
4.2	Article 1 : The Movement Disorders Clinical Practice	35
5	Analyse de survie : Développement d'un Nomogramme pronostique de décès pour les patients SCA	48
5.1	Résumé	48
5.1.1	Rationnel	48
5.1.2	Méthodes	48
5.1.3	Résultats	50
5.2	Article 2 : The Lancet Neurology	50
6	Analyse conjointe de la progression de l'ataxie et de la survie	68
6.1	Résumé	68
6.1.1	Rationnel	68
6.1.2	Méthodes	68
6.1.3	Résultat	71
6.2	Article 3 :	72
7	Discussion générale	102
7.1	Principaux résultats et leur impact	102
7.2	Limites de la thèse	104
7.3	Conclusion et Perspective	104
	Bibliographie	115

1 Introduction générale

1.1 Contexte

L'ataxie spinocérébelleuse (SCA) est une maladie neurodégénérative, héréditaire autosomique dominante, cliniquement et génétiquement hétérogène qui se manifeste par des troubles cérébelleux [Durr, 2010]. C'est une pathologie rare avec une prévalence estimée entre 1,2 et 3,7 pour 100 000 personnes [Schöls et al., 2004]. Néanmoins, plus de 40 génotypes en cause ont été identifiés, les plus communs étant le type 1 (ou SCA1), SCA2, SCA3 et SCA6. Ces ataxies sont dues à des expansions répétées du CAG codant pour des polyglutamines dans des protéines avec des fonctions différentes à chaque type [Durr, 2010].

L'absence de traitement qui ralentirait la progression de l'ataxie favorise l'augmentation du poids de sa morbidité et sa mortalité qui pèse de plus en plus sur la société, entraînant ainsi un sérieux problème de santé publique. L'ataxie débute entre 30 et 55 ans touchant ainsi les personnes en pleine activité professionnelle et progresse linéairement quel que soit le génotype entraînant à long terme une invalidité irréversible avec une altération de la qualité de vie puis une mort prématurée [Durr, 2010, Schöls et al., 1997]. Les dépenses engagées par les différents systèmes de santé pour sa prise en charge (test génétique, hospitalisation, réhabilitation, invalidité) se chiffrent à plusieurs milliers de dollars [López-Bastida et al., 2008].

Ces dernières années, plusieurs efforts ont été faits pour étudier et comprendre l'histoire naturelle de ces maladies. L'analyse de données longitudinales de la cohorte Européenne EUROSCA après 8 ans de suivi a permis de déterminer les taux de progression spécifiques à chaque génotype et d'identifier les facteurs influençant cette évolution. La progression de l'ataxie mesurée par le Scale for the Assessment and Rating of Ataxia (SARA) [Schmitz-Hubsch et al., 2006] est plus rapide chez les SCA1, intermédiaire chez les SCA2 et SCA3, et plus lente chez les SCA6. Certains facteurs influençant cette progression ont été identifiés notamment l'âge à l'inclusion, l'âge de début de la maladie, le score SARA à l'inclusion et la longueur d'expansion de l'allèle [Jacobi et al., 2015].

1.2 Problématique

L'approche statistique utilisée pour étudier la progression de l'ataxie au cours du temps en l'occurrence le modèle linéaire mixte [Laird and Ware, 1982], suppose que les patients d'un même gène suivent un profil moyen d'évolution du score SARA. Cependant, cette hypothèse est très forte et il n'est pas rare qu'une hétérogénéité soit suspectée. C'est-à-dire qu'au sein d'un même gène il existerait plusieurs sous-groupes d'évolution différente de l'ataxie. Des modèles appropriés

pouvant identifier ces sous-groupes ont été récemment proposés : les modèles à classes latentes [Muthén and Shedden, 1999, Verbeke and Molenberghs, 2009a] . Plusieurs études portant sur des patients atteints d’ataxie ou des modèles de souris transgéniques sur les SCA suggèrent que la progression de l’ataxie serait précédée d’une perte de poids [Saute et al., 2012, Rüb et al., 2006, Jafar-Nejad et al., 2011, Hübener et al., 2012, Boy et al., 2009]. Cette perte de poids s’expliquerait par l’atteinte d’autres structures du cerveau autre que le cervelet. Ainsi, l’identification de différents profils d’évolutions du poids utilisant ces méthodes et leur corrélation à la progression de la maladie permettrait de confirmer cette hypothèse, d’utiliser le poids pour le monitoring de la progression de l’ataxie et d’optimiser au mieux le design des futurs essais thérapeutiques.

Par ailleurs, bien que l’espérance de vie des patients atteints d’ataxie spinocérébelleuse soit considérablement réduite, seules quelques études limitées en termes de tailles d’échantillon et de suivi ont estimé la survie et identifié des facteurs influençant cette survie [Klockgether et al., 1998, Kieling et al., 2007, Almaguer-Mederos et al., 2013, Monin et al., 2015]. Cependant, il apparaît nécessaire de disposer des taux de survie spécifiques à chaque génotype et d’identifier les facteurs qui leurs sont associés à partir de données longitudinales de cohorte avec un long suivi.

Enfin, une fois ces taux de survie spécifique à chaque génotype estimés, il serait intéressant de les modéliser conjointement avec la progression de la maladie. Cette modélisation conjointe permettrait de mieux comprendre leur interdépendance, la prise en compte de leur corrélation et l’identification de facteurs pronostiques qui leur sont communs.

1.3 Objectifs

Face à tous ces défis, les objectifs de notre travail ont été :

- d’identifier des profils (ou trajectoires) d’évolution du poids,
- d’étudier la survie des patients atteints d’ataxie et,
- d’analyser conjointement la survie des patients et l’évolution de la maladie.

Le reste du document est organisé comme suit : le **chapitre 2** présente une revue de la littérature traitant des généralités sur l’ataxie spinocérébelleuse. D’une part nous décrirons la physiopathologie ainsi que les données épidémiologiques et cliniques de l’ataxie, et d’autre part la cohorte EUROSCA sur laquelle les différentes méthodes ont été appliquées. Un bref résumé des différentes méthodes utilisées au cours de cette thèse sera présenté dans le **chapitre 3**. Dans le **chapitre 4**, nous présentons les analyses de trajectoires notamment l’impact de la perte de poids sur la progression de l’ataxie. Le **chapitre 5** traite de l’analyse de survie. Dans le **chapitre 6**, nous présentons l’analyse conjointe entre la survie et la progression de l’ataxie. Enfin une conclusion résume l’apport de ce travail et expose quelques perspectives.

2 Etat des connaissances

2.1 Généralités

2.1.1 Définition

Les Ataxies Cérébelleuses Autosomiques Dominantes (ACAD) sont une cause rare d'ataxie cérébelleuse. Elles font parties d'un groupe de troubles neurologiques, cliniquement et génétiquement hétérogènes. Ce sont des maladies génétiques, neurodégénératives progressives qui se manifestent par des troubles cérébelleux résultant d'une démarche instable, d'une maladresse et d'une dysarthrie. Le syndrome cérébelleux est souvent associé à des signes pyramidaux ou extrapyramidaux comme une atteinte ophtalmologique ou une déficience cognitive [Durr, 2010, Schöls et al., 2004, Harding, 1983]. Le début de la maladie est le plus souvent entre la 3^{eme} ou la 4^{eme} décennie, mais parfois il survient à l'adolescence ou pendant la vieillesse [Klockgether, 2010]. L'atrophie du cervelet et du tronc cérébral sont les principales caractéristiques retrouvées à l'imagerie, mais d'autres structures peuvent être atteintes, conduisant ainsi à une large variété du phénotype [Klockgether, 2010]. Actuellement, plus de 40 génotypes en cause ont été identifiés. Selon le mécanisme mutationnel, ces génotypes se regroupent en 3 grandes catégories :

- **A : les Ataxies Spinocérébelleuses (SCA) à expansion de polyglutamine** qui résultent de répétitions excessives de trinuécléotides CAG codant la glutamine. Selon les gènes altérés, différents type sont définis. Ce sont les types 1 (ou SCA1), SCA2, SCA3, SCA6, SCA7, SCA17 et DRPLA (atrophie dentato-rubro-pallidoluysienne). Ces SCA partagent le même mécanisme mutationnel que les autres maladies à expansion de polyglutamine comme la maladie de Huntington [Group, 1993].
- **B : Les SCA à expansion non-codante** découvertes récemment, ces expansions non traduites peuvent entrainer une maladie par un mécanisme de gain de fonction, déclenché par l'accumulation de transcrits contenant des expansions répétées de certains acides aminés CTG pour SCA8, ATTCT pour SCA10, CAG pour SCA12 et TGGAA pour SCA31.
- **C : Les mutations conventionnelles** causées par des mécanismes de faux-sens (SCA5, SCA13, SCA14, SCA28 et SCA28), de duplication (SCA20), et enfin de faux-sens et délétion (SCA15/16).

2.1.2 Épidémiologie

Les ACAD sont des troubles neurologiques rares avec une prévalence estimée à moins de 4 pour 100000 personnes en Europe [Schöls et al., 2004]. Parmi elles, les SCA à expansion de polyglutamine sont les plus fréquentes avec une prédominance des SCA3 [Schöls et al., 2004]. Cependant, cette prévalence est sous-estimée car dans de nombreuses séries ne sont pris en compte que certains gènes représentants en moyenne que 44% [SD 25] des cas de ACAD (en fonction

de la région géographique d'origine) [Durr, 2010]. Ainsi, dans la plupart des études, seuls les tests génétiques pour les gènes les plus fréquents (SCA1, SCA2, SCA3, SCA6, SCA7 et SCA17) ont été effectués [Durr, 2010]. C'est le cas dans les cohortes d'histoire naturelle en Europe (EUROSCA) [Schmitz-Hubsch et al., 2008a] et aux USA (Clinical Research Consortium for Spinocerebellar Ataxias (CRC-SCA)) [Ashizawa et al., 2013] où seuls les SCA1, SCA2, SCA3 et SCA6 ont été inclus et suivis. Les proportions de patients dans chaque sous-type sont similaires entre les deux cohortes. Dans EUROSCA, parmi les 525 patients SCA inclus entre 2005 et 2006, et suivi depuis, les SCA2 sont les plus fréquents (31%), suivi des SCA3 (26%), les SCA1 (22%) et les SCA6 (20%). En revanche, dans le CRC-SCA initié en Avril 2010 avec 345 patients SCA inclus et suivis tous les 6 mois pendant 2 ans, ce sont les SCA3 qui sont les plus fréquents (40%), suivi des SCA2 (22%), les SCA6 (21%) et les SCA1 (17%). En France, sur une série de 826 cas index, les mutations les plus fréquentes étaient les SCA3 (20%), SCA2 (10%), SCA1 (8%), SCA7 (6%) et SCA6 (2%, figure 2.1) [Durr, 2010]. En plus, à cause de l'effet fondateur retrouvé pour certains génotypes notamment les SCA2 au Cuba (40/100000 cas) [Orozco et al., 1989] et les SCA3 au Portugal (714/100000 cas) [Sequeiros and Coutinho, 1993, Vale et al., 2010], la prévalence de ces ataxies varie différemment d'une région géographique à l'autre [Durr, 2010]. La fréquence relative des SCA3 est très élevée avec 69% des cas au Brésil [Teive et al., 2008], 58% au Portugal [Sequeiros and Coutinho, 1993, Vale et al., 2010], 49% en Chine [Tang et al., 2000], entre 26% et 63% au Japon [Basri et al., 2007, Maruyama et al., 2002, Shibata-Hamaguchi et al., 2009], 28% en Hollande [van de Warrenburg et al., 2002] et 42% en Allemagne [Schöls et al., 1997] ; faible en France (20%) [Durr, 2010], au Canada (24%) [Kraft et al., 2005] et aux USA (21%) [Moseley et al., 1998] ; et rare en Inde (3%) [Faruq et al., 2009], en Norvège (4%) [Erichsen et al., 2009] et en Italie (1%) [Brusco et al., 2004].

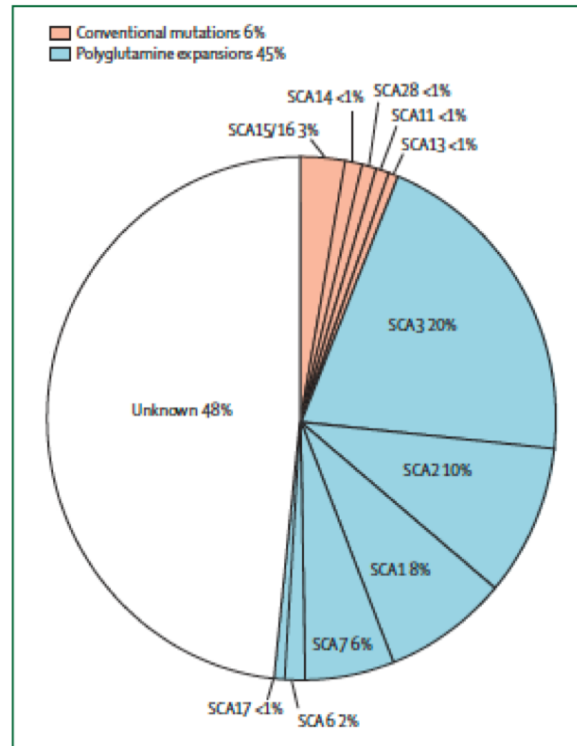
2.1.3 Mécanismes physiopathologiques

Plusieurs types de mécanismes physiopathologiques impliqués dans les SCA à expansion de polyglutamine ont été identifiés [Durr et al., 1996]. Les principaux sont résumés dans la table 2.1. Entre autres, on note la formation d'inclusion intranucléaire ou l'accumulation toxique d'agrégats incriminés dans les SCA1, SCA2 et SCA3 [Williams and Paulson, 2008, Todd and Paulson, 2009, Matilla-Dueñas et al., 2010] ; un dysfonctionnement dans la régulation de l'homéostasie du Ca^{2+} décrit dans les SCA2 et SCA6 [Liu et al., 2009, Kordasiewicz and Gomez, 2007] ; un défaut dans la transmission du glutamate impliqué dans les SCA1 et SCA3 [Matilla-Dueñas et al., 2010] ; des altérations de l'ARN dans les SCA1 et SCA2 [Todd and Paulson, 2009, Matilla-Dueñas et al., 2010] ; et enfin une dysfonction mitochondriale commune à tous les groupes [Simon et al., 2007].

TABLE 2.1 – Mécanismes physiopathologiques impliqués dans les SCA à expansion de glutamine

Mécanismes physiopathologiques	Types de SCAs à expansion de polyglutamine
Accumulation toxique d'agrégats et inclusion intranucléaire	SCA1, SCA2, SCA3
Inférence de transcription	SCA3
Homéostasie du Ca^{2+}	SCA2, SCA6
Transmission du glutamate	SCA1, SCA3
Dysfonction mitochondriale	SCA1, SCA2, SCA3, SCA6
Altérations de l'ARN	SCA1, SCA2

FIGURE 2.1 – Prévalence relative des sous types de SCA parmi les ataxies cérébelleuses autosomiques dominantes (données françaises de 826 cas index)[Durr, 2010]



2.1.4 Signes cliniques

Les SCA à expansion de polyglutamine ont en commun le même mécanisme mutationnel résultant de la répétition du CAG. Pour devenir pathologique, le nombre de répétitions du CAG doit atteindre un certain seuil (37-40) variant d'un génotype à l'autre [?]. Au-dessus de ce seuil, apparaît une dysfonction neurologique diffuse entraînant des troubles neurologiques allant jusqu'au décès par défaillance du tronc cérébral [Durr, 2010, Schöls et al., 2004]. Les corrélations entre les phénotypes et les génotypes ont montré que la différence dans la taille de la répétition contribue à la sévérité et à la progression de la maladie. Ainsi la taille de répétition est négativement corrélée à l'âge de début de la maladie : plus la longueur de la répétition est grande plus l'âge de début de la maladie est précoce [Yamada et al., 2007]. En moyenne les premiers symptômes chez les patients porteurs d'un SCA1 ou SCA2 surviennent entre 30 et 40 ans, autour de 40 ans pour les SCA3 et plus tard entre 50 et 60 ans pour les SCA6. Quel que soit le génotype, l'ataxie de type cérébelleuse est le symptôme majeur [Schöls et al., 2004]. Une étude évaluant la précocité des symptômes autres que les troubles de la marche qui, en général apparaissent en premier dans 2/3 des cas, montre que la vision double, la dysarthrie, les épisodes de vertiges, l'incoordination des membres et des anomalies oculomotrices cérébelleuses ont précédé l'ataxie chez 4% des patients [Globas et al., 2008]. Les principaux signes caractéristiques à chaque génotype ainsi que le gène en cause sont résumés dans la table 2.2 [Klockgether, 2008].

TABLE 2.2 – Génétiques moléculaires et phénotypes cliniques des SCA

Génotype	Mutation	Gène	Phénotype clinique
SCA1	Expansion du CAG	ATXN1	Ataxie, signes pyramidaux, dysphagie, syndrome des jambes sans repos
SCA2	Expansion du CAG	ATXN2	Ataxie, signes ophtalmologiques (saccades faibles), neuropathies, syndrome des jambes sans repos
SCA3	Expansion du CAG	ATXN3	Ataxie, signes pyramidaux, signes ophtalmologiques, neuropathies, dystonie, syndrome des jambes sans repos
SCA6	Expansion du CAG	CACNA1A	Presque pure ataxie cérébelleuse

2.1.5 Signes radiologiques

L'atrophie du cervelet est le signe radiologique majeur des ataxies cérébelleuses, mais dans les SCA à expansion de polyglutamine, l'atrophie du vermis est le plus souvent le seul changement notable dans le cervelet [Durr, 2010]. L'atrophie pontique prédomine spécialement dans les SCA1 et SCA2 [Guerrini et al., 2004]. En plus, il a été montré que les dysfonctions cliniques sont bien corrélées avec l'atrophie du pont dans les SCA1, de la totalité du tronc cérébral dans les SCA3 et du cervelet dans les SCA6 (figure 2.2) [Schulz et al., 2010]. Globalement, tous les patients SCA présentent des lésions du tronc cérébral [Yamada et al., 2007]. Les cellules de Purkinje sont affectées sauf dans les SCA3, et la majorité des patients ont une perte neuronale dans le ganglion basal. Cette perte est plus répandue dans les SCA1, avec une prédominance de l'atrophie des nerfs crâniens et la perte des cellules de Purkinje dans le cervelet [Yamada et al., 2007]. Dans les SCA2, l'atrophie du pont, de l'olive inférieur, de la substance grise et du cervelet sont les signes clés, mais également une perte cellulaire a aussi été reportée. Les lésions du ganglion basal, des colonnes inter-médio-latérale et de Clarke sont plus sévères dans les SCA3 [Yamada et al., 2007, Durr et al., 1996].

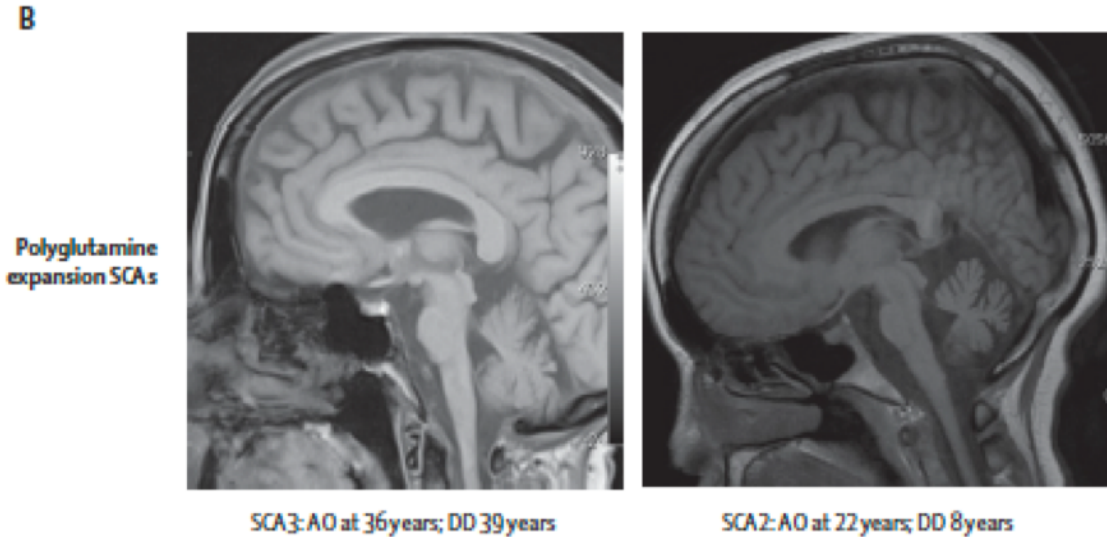
2.1.6 Diagnostic

Bien que chaque SCA possède un certain nombre de caractéristiques qui nous permettent de définir un phénotype clinique typique, il existe une grande intersection phénotypique entre les différents SCA. En raison de cette intersection et de la grande variabilité observée au sein des familles, il est presque impossible de poser un diagnostic fiable simplement sur des bases cliniques [Klockgether, 2008]. Le diagnostic de certitude repose donc sur la positivité des tests génétiques spécifiques à chaque gène. Cependant, ces tests nécessitent une haute technologie et sont extrêmement coûteux [López-Bastida et al., 2008] pouvant peut être expliquer le faible nombre de cas d'ataxie spinocérébelleuse rapporté dans les pays en développement.

2.2 Cohorte EUROSICA

Les données sur lesquelles ont été appliquées les différentes méthodes au cours de cette thèse proviennent de l'étude de cohorte Européenne, multicentrique (17 centres), longitudinale, et prospective : EUROSICA. Cette étude a été initiée en 2005 avec pour but de caractériser l'évolution de l'histoire naturelle des ataxies spinocérébelleuses et d'identifier ses principaux facteurs de risque.

FIGURE 2.2 – Atteinte du cervelet chez des patients SCA2 (à droite) et SCA3 (à gauche)[Durr, 2010] : coupe sagittale de l’Imagerie à Raisonance Magnétique (IRM) montrant l’implication du tronc cérébral avec une atteinte relativement mineure du cervelet dans les SCA à expansion de polyglutamine (SCA2, SCA3) chez des patients avec une durée de la maladie comparable; SCA = ataxie spinocérébelleuse. AO = âge au début de la maladie. DD = durée de la maladie.



Entre le 01 Juillet 2005 et le 31 Aout 2006, 525 patients ont été identifiés et recrutés à partir d’un registre électronique contenant les données de tous les patients atteints d’ataxie spinocérébelleuse qui avaient été en contact avec l’un des 17 centres d’études. Ces patients étaient âgés d’au moins 18 ans et présentaient un test génétique positif pour les SCA1 (n=117), SCA2 (n=162), SCA3 (n=139) ou SCA6 (n=107), et une ataxie progressive ou inexpliquée. L’étude a été planifiée selon un protocole bien défini ; les patients ont été vu à l’inclusion et suivis par des visites annuelles pendant trois ans (± 3 mois). Après, la cohorte est entrée dans une phase extensive entraînant ainsi des intervalles de visites irrégulières. A chaque visite, une batterie de tests neurologiques a été administrée aux patients (Scale for the Assessment and Rating of Ataxia (SARA) [Schmitz-Hubsch et al., 2006], Inventory of Non-Ataxia Signs (INAS) [Jacobi et al., 2013a], SCA Functional Index (SCAFI) [Schmitz-Hubsch et al., 2008b], Patient Health Questionnaire-9 score (PHQ-9) [Spitzer et al., 1999], Visual Analogue Scale (VAS) of the EuroQol five dimensions questionnaire (EQ-5D) [EuroQol Group, 1990] et functional assessment (part IV) of the Unified Huntington’s Disease Rating Scale (UHDRS)) [Group, 1996], ainsi que d’autres informations sur leur état de santé telles que : l’âge, le sexe, le centre, le génotype, la famille, la taille des expansions de CAG sur chacun des allèles (CAG long et court), l’âge de début de la maladie, la durée de la maladie, le poids, la taille et le statut vital.

2.2.1 Mesures de la sévérité de l’ataxie

Deux échelles ont été analysées au cours de cette thèse : le score SARA qui mesure la sévérité de l’ataxie et le score semi-quantitatif INAS qui évalue les symptômes neurologiques autres que l’ataxie.

Le Scale for the Assessment and Rating of Ataxia (SARA)

Le score SARA a été conçu lors d’une réunion tenue en Février 2004 à Bonn, en Allemagne par un groupe de neurologues Européens. Les items du score ont été sélectionnés en fonction de la spécificité de l’ataxie et de la possibilité de standardiser les procédures de test et de notation. De plus, l’évaluation devait être proche de l’examen neurologique standard et ne devait pas nécessiter d’équipements techniques. Brièvement, SARA comprend huit items qui sont ; **1 : démarche** (score de 0 à 8), **2 : position debout** (score de 0 à 6), **3 : position assise** (score de 0 à 4), **4 : trouble de la parole** (score de 0 à 6), **5 : poursuite du doigt** (score de 0 à 4), **6 : test de nez-doigt** (score 0 à 4), **7 : mouvements de la main alternant rapidement** (score 0 à 4), **8 : glisser du talon sur le tibia** (score 0 à 4). Les fonctions cinétiques des membres (items 5 à 8) sont évaluées indépendamment pour les deux côtés, et la moyenne arithmétique des deux côtés est incluse dans le calcul du score total de SARA. Ces items évaluent à la fois l’ataxie axiale ou centrale (items de 1 à 4) et l’ataxie périphérique (items de 5 à 8) donnant un score total allant de 0 (absence d’ataxie) à 40 (ataxie sévère). [Schmitz-Hubsch et al., 2006] ont montré que ce score était linéaire, unidimensionnel avec de très bonnes propriétés métriques : consistance interne avec le coefficient α de Cronbach à 0.94 ; une fiabilité inter-évaluateur de 0.97 (items pris séparément ou score total), et une fiabilité intra-évaluateur (test-retest) de 0.98. En plus, la sensibilité au changement et la reproductibilité du SARA ont été évaluées en utilisant l’impression globale du patient comme mesure d’encrage regroupé en ‘mauvais’, ‘stable’ et ‘meilleur’ statut de la maladie par rapport à l’inclusion. Ces données ont montré que la sensibilité au changement et la reproductibilité du SARA étaient bonnes [Schmitz-Hubsch et al., 2010]. Depuis SARA a été validé et, est utilisé comme critère de jugement principal pour mesurer la progression de la sévérité de l’ataxie dans de nombreuses études épidémiologiques [Jacobi et al., 2011, Ashizawa et al., 2013]. Cependant, son utilisation dans les futurs essais cliniques ne fait pas encore l’unanimité et reste une véritable question de recherche.

Le Inventory of Non-Ataxia Signs (INAS)

INAS a été développé pour évaluer les autres symptômes neurologiques non-ataxiques. Il se compose de 30 items, dont chacun est lié à l’un des 16 symptômes ou syndromes suivant : **aréflexie, hyper-réflexie, réponse plantaire extenseur, spasticité, parésie, amyotrophie, fasciculations, myoclonie, rigidité, chorée, dystonie, tremblement au repos, symptômes sensoriels, signes oculomoteurs du tronc cérébral (ophtalmo-parésie horizontale et verticale, ralentissement des saccades), dysfonction urinaire et troubles cognitifs**. En affectant 1 ou 0 à la présence ou l’absence de chacun de ces signes, on obtient un nombre total de signes non ataxiques compris entre 0 et 16. Par ailleurs, ces symptômes non ataxiques ont été étendus à la dysphagie et la double vision. Comme le SARA, le score semi-quantitatif INAS a été validé de façon rigoureuse et présente aussi de très bonnes propriétés métriques avec une fiabilité inter-évaluateur de 0.88, et une bonne reproductibilité [Jacobi et al., 2013a]. En revanche, sa sensibilité au changement est moins bonne [Schmitz-Hubsch et al., 2010]. Cependant, l’ensemble des études concluent que INAS est une mesure valide de l’implication extra-cérébelleuse dans les troubles de l’ataxie progressive, et peut être utilisé comme complément au SARA mais jamais comme critère de jugement principal dans les futurs essais thérapeutiques [Jacobi et al., 2013a].

2.2.2 Etude de l’histoire naturelle de l’ataxie

Progression à long terme de l’ataxie mesuré par le SARA

Les analyses de données longitudinales de la cohorte EUROSCA après 2 ans [Jacobi et al., 2011] et 8 ans [Jacobi et al., 2015] de suivi ont permis de déterminer les taux de progression spécifiques à chaque génotype et d’identifier les facteurs influençant cette évolution. [Jacobi et al., 2015] ont montré que l’ataxie progressive est le symptôme majeur de ces troubles neurologiques et augmente de façon linéaire au cours du temps, avec des taux de progression différents pour chaque génotype (Figure 3). La progression de l’ataxie est plus rapide chez les SCA1 (2.1 ± 0.12 point SARA par an), intermédiaire chez les SCA2 (1.49 ± 0.07) et SCA3 (1.56 ± 0.08), et plus lente chez les SCA6 (0.80 ± 0.09). Les facteurs liés à cette progression sont l’âge avancé et la longueur de l’expansion du CAG long chez les SCA1, l’âge précoce de début des symptômes chez les SCA2 et le score SARA à l’inclusion à la fois chez les SCA2 et SCA6. Aucun facteur n’a été retrouvé, pour les patients SCA3. Par ailleurs, aux USA les taux de progression estimés à partir de la cohorte CRC-SCA [Ashizawa et al., 2013] sont comparables aux données EUROSCA [Jacobi et al., 2015].

Progression à long terme des symptômes non-ataxique mesuré par INAS

La progression des symptômes neurologiques non-ataxiques mesurés par le score semi-quantitatif INAS n’est linéaire que chez les patients SCA6, avec une progression plus lente par rapport aux autres SCA. Pour les SCA1, SCA2 et SCA3, la progression de l’INAS est non-linéaire (fonction quadratique du temps) avec des taux de progression similaires. Les facteurs associés à cette progression sont la précocité de l’âge de début de la maladie pour les patients SCA1 et SCA2, moins de symptômes neurologiques non-ataxiques à l’inclusion pour les SCA1, SCA2 et SCA6, et le sexe pour les SCA3 avec une progression plus rapide chez les femmes [Jacobi et al., 2015].

FIGURE 2.3 – Progression du score SARA chez les patients SCA1 (A), SCA2 (B), SCA3 (C) et SCA6 (D)[Jacobi et al., 2015] : Les données sont représentées par la moyenne (IC à 95%). Les points rouges, noirs et bleus indiquent les valeurs moyennes. Les lignes noires continues indiquent les valeurs observées de tous les patients d'un génotype. Dans les SCA1 et SCA2, la courbe bleue représente les valeurs observées par les participants avec un suivi pouvant aller jusqu'à 3 ans (± 3 mois) et les participants à la courbe rouge avec un suivi plus long (> 3 ans). Comme il n'y avait pas de différence entre SCA3 et SCA6 entre les patients avec un suivi plus court et plus long, seules les valeurs observées pour tous les patients sont présentées pour ces génotypes. Les courbes en pointillés indiquent la progression estimée ou prédite par le Pattern-Mixture Model (PMM) tenant compte des données manquantes ; SARA = échelle d'évaluation et de notation de l'ataxie et SCA = ataxie spinocérébelleuse.

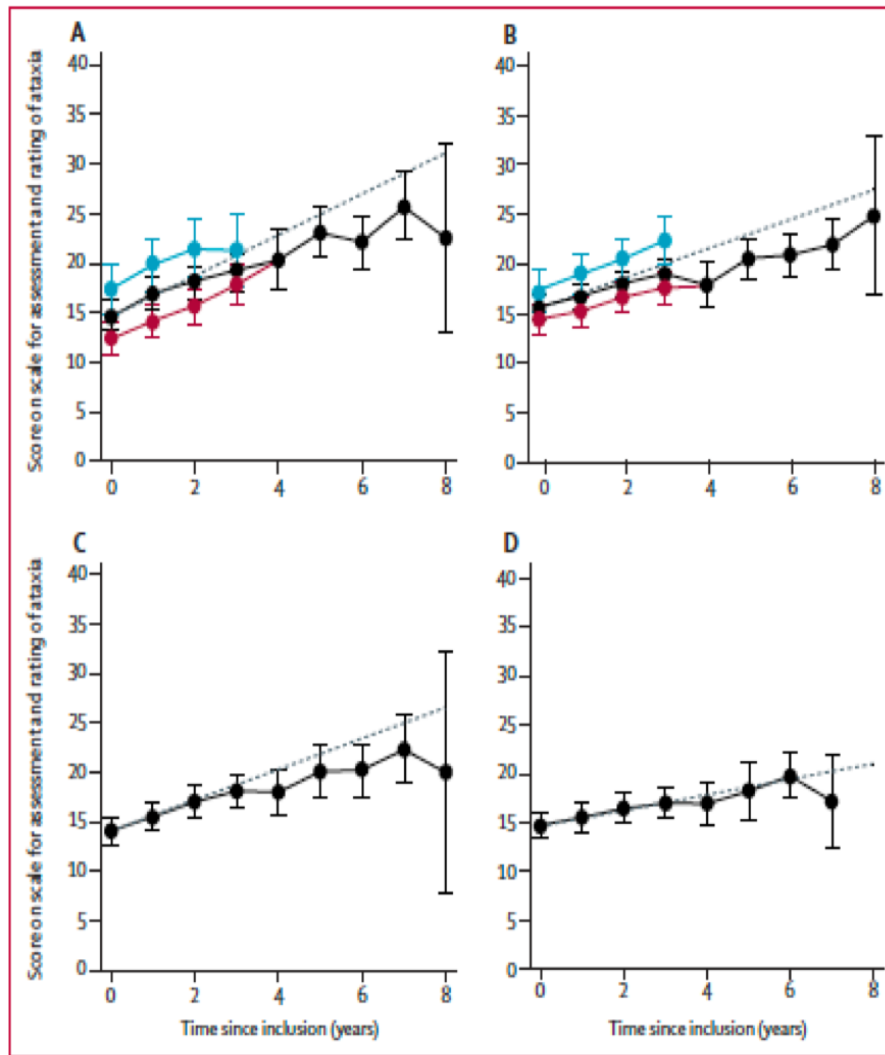
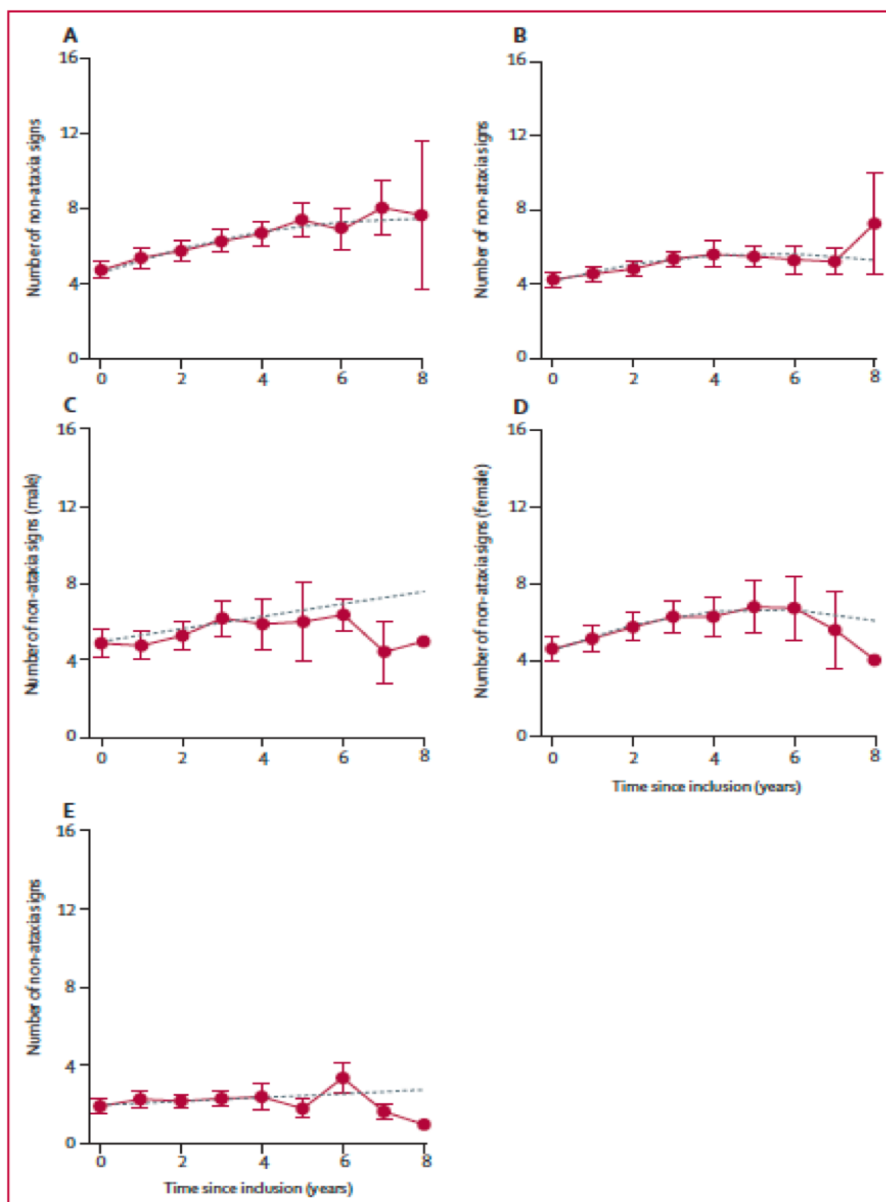


FIGURE 2.4 – Progression du score semi-quantitatif INAS chez les patients SCA1 (A), SCA2 (B), les hommes SCA3 (C), les femmes SCA3 (D) et SCA6 (E)[Jacobi et al., 2015] : Les données sont représentées par la moyenne (IC à 95%). Comme les résultats montraient une interaction entre le sexe et l'augmentation des signes de non-ataxique chez les patients SCA3 uniquement, le graphique pour les patients SCA3 est représenté séparément par sexe. Les lignes rouges continues indiquent les valeurs observées. Les courbes en pointillés montrent la progression estimée ou prédite par le modèle linéaire mixte standard. Il manque des barres d'erreur sur le délai de 8 ans sur les panels C et E car trop peu de patients ont été suivis lors de cette visite ; SCA = ataxie spinocérébelleuse.



3 Méthodologie utilisée au cours de la thèse

Dans ce chapitre, nous exposons un bref résumé des modèles proposés dans la littérature pour traiter les données longitudinales (mesures répétées de variables quantitatives) et les données de survie (délai de survenue d'évènements censurés à droite). Nous décrirons la méthodologie des modèles mixtes, puis leur extension aux modèles à classe latente, et enfin les modèles conjoints pour données longitudinales et de survie.

3.1 Méthodes pour les données longitudinales

3.1.1 Modèles linéaires mixtes

Le modèle linéaire mixte a été proposé pour analyser les données d'une variable quantitative gaussienne mesurée à plusieurs reprises, à différents temps, sur les mêmes patients, tout en prenant en compte la corrélation intra-patient [Laird and Ware, 1982].

Spécification du modèle linéaire mixte

Soit Y_{ij} l'observation de la variable quantitative gaussienne Y pour le patient $i, i = 1, \dots, N$ au temps $t_{ij}, j = 1, \dots, n_i$. Le modèle linéaire standard s'écrit :

$$Y_{ij} = X_{ij}^T \beta + Z_{ij}^T b_i + \varepsilon_{ij} \quad (3.1)$$

où $b_i \sim N(0, B)$ sont des q -vecteurs d'effets aléatoires distribués identiquement et indépendamment suivant une loi normale de moyenne 0 et de matrice de variance covariance B ; $\varepsilon_{ij} \sim N(0, \sigma_e^2)$ les erreurs de mesures gaussiennes indépendantes des effets aléatoires b_i ; X_{ij} le vecteur de variables explicatives de dimension p pour la mesure j du patient i , β le p -vecteur des effets fixes incluant généralement l'intercept global β_0 ; Z_{ij} un sous-vecteur de X_{ij} de dimension $q (q \leq p)$ incluant différentes fonctions du temps (linéaire, quadratique, cubique, spline, etc...).

Estimation

La vraisemblance du modèle linéaire mixte s'écrit :

$$\mathcal{L}(\beta, \theta) = -\frac{1}{2} \sum_{i=1}^N (n_i \log(2\pi) + \log |V_i| + (Y_i - X_i \beta)^T V_i^{-1} (Y_i - X_i \beta)) \quad (3.2)$$

où θ est le vecteur des paramètres de covariance intervenant dans V_i ; et $|V_i|$ est le déterminant de V_i . La vraisemblance est maximisée en (β, θ) suivant l'équation du score :

$$\frac{\partial \mathcal{L}(\beta, \theta)}{\partial \beta} = \sum_{i=1}^N X_i^T V_i^{-1} (Y_i - X_i \beta) = 0 \quad (3.3)$$

Lorsque les paramètres θ sont connus, β est estimé par :

$$\hat{\beta} = \left(\sum_{i=1}^N X_i^T V_i^{-1} X_i \right)^{-1} \left(\sum_{i=1}^N X_i^T V_i^{-1} Y_i \right) \quad (3.4)$$

Ainsi θ est estimé en maximisant $\mathcal{L}(\beta, \theta)$ par une procédure itérative telle que l'algorithme de Newton-Raphson ou de Quasi-newton le plus souvent.

3.1.2 Modèles mixtes à classes latentes

Le modèle mixte à classe latente a été initialement défini par Verbeke et Lesaffre pour tester l'écart à la normalité des effets aléatoires [Verbeke and Molenberghs, 2009a] et puis par [Muthén and Shedden, 2001] comme une extension de la théorie des modèles mixtes standards à l'étude des populations hétérogènes.

Formulation du modèle

Le modèle linéaire mixte à classe latente suppose qu'il existe un nombre fini G de sous-populations homogènes, définissant G classes latentes où chaque patient $i, i = 1, \dots, N$ appartient à une seule classe $g, g = 1, \dots, G$. Ils comprennent deux sous-modèles [Commenges and Jacqmin-Gadda, 2016] :

- **(a) une régression logistique multinomiale** qui définit pour chaque patient i , la probabilité qu'il appartienne à la classe g

$$\pi_{ig} = P(c_i = g | X_{1i}) = \frac{\exp(\zeta_{0g} + X_{1i}^T \zeta_{1g})}{\sum_{l=1}^G \exp(\zeta_{0l} + X_{1i}^T \zeta_{1l})} \quad (3.5)$$

où c_i est la variable latente discrète, égale à g si le patient i appartient à la classe g ; Le vecteur X_{1i} inclut les variables explicatives possiblement dépendantes du temps. Pour que le modèle soit identifiable, on suppose que $\zeta_{0G} = 0$ et $\zeta_{1G} = 0$; et

- **(b) un modèle linéaire mixte** qui définit la trajectoire du marqueur Y conditionnel à chaque classe g :

$$Y_{ij} = X_{2ij}^T \beta_g + Z_{ij}^T b_{ig} + \varepsilon_{ij} \quad (3.6)$$

où β_g est le vecteur de paramètre de régression spécifique à chaque classe g associé au vecteur de variables explicatives X_{2ij} , Z_{ij} un sous-vecteur de X_{2ij} ; $b_{ig} \sim N(0, B_g)$ les effets aléatoires avec $B_g = B, \forall g \in [1, G]$ la matrice de variance-covariance que nous avons supposé commune aux classes; les erreurs de mesures $\varepsilon_{ij} \sim N(0, \sigma_e^2)$ sont indépendantes entre elles et indépendantes des effets aléatoires.

Estimation

L'estimation des paramètres du modèle linéaire mixte à classe latente se fait par maximum de vraisemblance à nombre de classe latente G fixé [Commenges and Jacqmin-Gadda, 2016]. La

distribution de Y_i marginale par rapport aux effets aléatoires et conditionnelle aux classes latentes est utilisée pour l'estimation :

$$[Y_i|c_i = g] \sim N(X_{2i}\beta + Z_i b_{ig}, Z_i B_i Z_i^T + \Sigma_i) \quad (3.7)$$

où X_{2i} et Z_i sont les matrices ayant respectivement comme vecteur ligne X_{2ij}^T et Z_{ij}^T avec $j = 1, \dots, n_i$. En notant θ le vecteur de paramètres intervenant dans les équations (3.5) et (3.6), la vraisemblance se définit donc :

$$\mathcal{L}(\theta) = \prod_{i=1}^N \left(\sum_{g=1}^G P(c_i = g | X_{1i}, \theta) \times \phi_{ig}(Y_i | c_i = g; X_{2i}, Z_i, \theta) \right) \quad (3.8)$$

où ϕ_{ig} est la densité de la loi normale multivariée d'espérance $X_{2i}\beta + Z_i b_{ig}$ et de variance $Z_i B_i Z_i^T + \Sigma_i$. Dans cette thèse, nous avons utilisé le package R *lcmm* développé par Proust-Lima [Proust-Lima et al., 2017] qui implémente l'algorithme de Marquardt modifié pour maximiser la log-vraisemblance $L(\theta) = \log(\mathcal{L}(\theta))$. A cause de la possibilité de convergence vers un maximum local, il est conseillé de relancer la procédure d'estimation en faisant varier les valeurs initiales pour assurer une convergence vers un maximum global [Commenges and Jacqmin-Gadda, 2016] .

Discrimination du modèle

Après l'estimation des paramètres, les patients peuvent être affectés dans la classe latente à laquelle, le patient a la plus grande probabilité d'appartenir $\argmax_g(\hat{\pi}_{ig}, g = 1, \dots, G)$ avec :

$$\hat{\pi}_{ig}^Y = P(c_i = g | X_i, Y_i, \hat{\theta}) = \frac{P(c_i = g | X_{1i}, \hat{\theta}) \phi_{ig}(Y_i | c_i = g; X_{2i}, Z_i, \hat{\theta})}{\sum_{m=1}^G P(c_i = m | X_{2i}, \hat{\theta}) \phi_{im}(Y_i | c_i = m; X_{2i}, Z_i, \hat{\theta})} \quad (3.9)$$

A partir de ces probabilités, on peut calculer une matrice de classification avec P_{lm} éléments représentant la probabilité moyenne *a posteriori* $\hat{\pi}_{il}^Y$ d'appartenir à chacune des classes l au sein de chaque classe *a posteriori* g . La classification est discriminante si $\hat{\pi}_{il}^Y$ est proche de 1 pour tout $l = g$ et $\hat{\pi}_{il}^Y$ proche de 0 si $l \neq g$ (Tableau 3.1) [Commenges and Jacqmin-Gadda, 2016]. Plusieurs autres critères peuvent être utilisés pour évaluer la discrimination [Commenges and Jacqmin-Gadda, 2016] :

- la proportion de patients ayant une probabilité *a posteriori* maximale au dessus d'un certain seuil (0.8 par exemple).
- l'entropie qui est d'autant plus proche de 1 que la classification est bonne. Elle vaut : $1 - \frac{En}{N \ln G}$

$$\text{avec } En = - \sum_{i=1}^N \sum_{g=1}^G \hat{\pi}_{ig}^Y \log(\hat{\pi}_{ig}^Y)$$

Sélection du nombre de classes latentes

Pour sélectionner le nombre optimal de classes latentes, plusieurs critères ont été proposés [Commenges and Jacqmin-Gadda, 2016] :

- le plus couramment utilisé est le Bayesian Information Criterion (BIC) : $BIC = -2\mathcal{L}(\theta) + p \log(N)$ où p est le nombre de paramètres estimés (dimension de θ).
- le Integrate Classification Likelihood (ICL) avec une approximation du BIC [Han et al., 2007a] qui, en plus de l'ajustement, prend en compte la classification *a posteriori* obtenue par le modèle : $ICL - BIC = BIC + 2 \times En$.

TABLE 3.1 – Tableau de classification *a posteriori* :

Classe Finale	#	Probabilité moyenne d'appartenir à chaque classe					
		1	...	g	...	G	
1	N_1	$\frac{1}{N_1} \sum_{i=1}^{N_1} \hat{\pi}_{i1}$	...	$\frac{1}{N_1} \sum_{i=1}^{N_1} \hat{\pi}_{ig}$	...	$\frac{1}{N_1} \sum_{i=1}^{N_1} \hat{\pi}_{iG}$	
$\vdots$		$\vdots$	$\ddots$			$\vdots$	
g	N_g	$\frac{1}{N_g} \sum_{i=1}^{N_g} \hat{\pi}_{i1}$	...	$\frac{1}{N_g} \sum_{i=1}^{N_g} \hat{\pi}_{ig}$	...	$\frac{1}{N_g} \sum_{i=1}^{N_g} \hat{\pi}_{iG}$	
$\vdots$		$\vdots$			$\ddots$	$\vdots$	
G	N_G	$\frac{1}{N_G} \sum_{i=1}^{N_G} \hat{\pi}_{i1}$	...	$\frac{1}{N_G} \sum_{i=1}^{N_G} \hat{\pi}_{ig}$	...	$\frac{1}{N_G} \sum_{i=1}^{N_G} \hat{\pi}_{iG}$	

Adéquation du modèle

L'adéquation du modèle aux données [Commenges and Jacqmin-Gadda, 2016] peut être évaluée de plusieurs manières :

- La comparaison entre les observations du marqueur et les prédictions marginales $\hat{Y}_{ijg}^{(M)}$ ou conditionnelles aux effets aléatoires $\hat{Y}_{ijg}^{(SS)}$. Ces prédictions sont toujours spécifiques à la classe g . Les prédictions marginales sont pondérées par la probabilité marginale d'appartenir aux classes, tandis que les prédictions conditionnelles sont pondérées par la probabilité *a posteriori* d'appartenir aux classes.

$$\hat{Y}_{ijg}^{(M)} = X_{2ij}^T \hat{\beta}$$

et

$$\hat{Y}_{ijg}^{(SS)} = X_{2ij}^T \hat{\beta} + Z_{ij}^T \hat{b}_{ig}^*$$

où $\hat{b}_{ig}^* = B_g Z_i^T V_{ig}^{-1} (Y_i - X_{2i} \hat{\beta})$ est l'estimateur empirique bayésien des effets aléatoires centrés dans la classe g .

Ces prédictions sont résumées soit en les moyennant sur les classes latentes pour obtenir une unique prédiction par patient :

$$\hat{Y}_{ij}^{(\cdot)} = \sum_{g=1}^G \hat{\pi}_{ig} \hat{Y}_{ijg}^{(\cdot)}$$

soit en les moyennant sur les sujets pour obtenir, à un temps de mesure donnée t , une prédiction par classe :

$$\hat{Y}^{(\cdot)}(t)_g = \sum_{i=1}^{N(t)} \hat{\pi}_{ig} \hat{Y}_{ijg}^{(\cdot)}$$

où $N(t)$ est le nombre de patients avec des mesures répétées au temps t .

Les résidus correspondants peuvent être déduits : $\hat{R}_{ij}^{(\cdot)} = Y_{ij} - \hat{Y}_{ij}^{(\cdot)}$.

- La qualité de la discrimination évaluée par la probabilité moyenne *a posteriori* d'appartenir aux classes décrit à l'équation (3.9). Une discrimination parfaite contiendrait des 1 en diagonale et des 0 ailleurs (Tableau 3.1). Cependant, une probabilité moyenne *a posteriori* autour de 0.6 indique une classification ambiguë, c'est-à-dire que les classes ne sont pas nettement différenciables et que les patients ne sont pas clairement associés à l'une des ces classes [Commenges and Jacqmin-Gadda, 2016].

3.2 Modèles conjoints

Les modèles conjoints ont été proposés pour étudier la relation entre les mesures répétées d'un marqueur et le délai de survenue d'un évènement [Lin et al., 2002]. Ils sont utiles pour l'étude de l'association entre les mesures répétées d'un marqueur et le processus de survie, ou pour décrire l'évolution au cours du temps d'un marqueur lorsque le suivi est interrompu par la survenue d'un évènement (une sortie informative), ou prédire le délai de survenue d'un évènement en fonction des mesures répétées d'un marqueur [Commenges and Jacqmin-Gadda, 2016]. Le principe est de combiner par des structures latentes un modèle mixte pour l'évolution du marqueur et un modèle de survie pour le risque d'évènement. Selon, la structure latente de liaison pour capturer la dépendance entre le marqueur et le délai d'évènement, deux approches ont été proposées : le modèle conjoint à effets aléatoires partagés [Rizopoulos, 2012] où la dépendance est capturée par les effets aléatoires et le modèle conjoint à classe latente [Proust-Lima et al., 2014, Commenges and Jacqmin-Gadda, 2016, Lin et al., 2002] où ce sont les classes latentes qui capturent toute la dépendance entre les deux processus.

3.2.1 Modèle conjoint à effets aléatoires partagés

Le modèle conjoint pour un marqueur longitudinal gaussien et un temps d'évènement a été proposé par Wulfsohn et Tsiatis [Rizopoulos, 2012] dans lequel leur relation est capturée par les effets aléatoires qui capturent également la corrélation des mesures répétées du marqueur. Ainsi, conditionnellement aux effets aléatoires, les deux variables (marqueur et temps d'évènement) sont indépendantes.

Spécification du modèle conjoint à effets aléatoires partagés

le modèle conjoint à effets aléatoires comprends deux sous-modèles :

Un modèle linéaire mixte pour décrire la trajectoire du marqueur, conditionnellement aux effets aléatoires :

$$Y_{ij} = \tilde{Y}_i(t_{ij}) + \varepsilon_{ij} = X_{ij}^T \beta + Z_{ij}^T b_i + \varepsilon_{ij} \quad (3.10)$$

où $b_i \sim N(0, B)$ sont les d'effets aléatoires spécifiques aux patients, avec une matrice de variance covariance B . La valeur observée Y_{ij} au temps t_{ij} est la somme de la vraie valeur du marqueur $\tilde{Y}_i(t_{ij})$ et de l'erreur de mesure gaussienne $\varepsilon_{ij} \sim N(0, \sigma_e^2)$ supposées indépendantes entre elles et des effets aléatoires.

Un modèle à risque proportionnel pour modéliser le délai de survenue de l'évènement, conditionnel aux effets aléatoires :

$$h_i(t|b_i, \beta, \alpha, \alpha^{(a)}) = h_0(t) \exp(W_i^T \alpha + \kappa(b_i, \beta, t_{ij}, X_{ij}, Z_{ij})^T \alpha^{(a)}) \quad (3.11)$$

où $h_0(t)$ est le risque de base qui, contrairement au modèle de Cox doit être spécifié pour éviter de sous-estimer les écarts-types des paramètres estimés [Rizopoulos, 2012]. Le plus souvent, une distribution paramétrique de type Exponentiel, Weibull ou Gamma est utilisée. α est le vecteur de paramètres de régression associés au vecteur de variables explicatives W_i ; $\kappa(\cdot)$ est la fonction de paramétrisation qui définit la structure de dépendance des deux processus (longitudinal et survie). Elle est spécifiée *a priori*, et dépend des effets aléatoires et possiblement du temps, des variables et des paramètres du modèle mixte. Plusieurs formulations de cette fonction ont été proposées, mais nous y reviendrons dans le chapitre 7 sur les modèles conjoints. $\alpha^{(a)}$ est le vecteur de paramètre associé à $\kappa(\cdot)$, qui quantifie la force de l'association entre les deux processus.

Estimation du modèle

Soit T_i le délai de survenue de l'évènement δ_i l'indicateur de l'évènement qui vaut 1 si le patient i a eu l'évènement et 0 sinon. Selon l'hypothèse d'indépendance conditionnelle entre Y_i et (T_i, δ_i) conditionnellement aux effets aléatoires b_i , la contribution individuelle à la distribution conjointe s'écrit :

$$f_{Y_i, T_i, \delta_i}(Y_i, T_i, \delta_i, \theta) = \int_{R^{n_b}} f_{Y_i|b_i}(Y_i|b_i; \theta) f_{T_i, \delta_i|b_i}(T_i, \delta_i|b_i; \theta) f_{b_i}(b_i; \theta) db_i$$

avec R^{n_b} la dimension des effets aléatoires, θ inclut l'ensemble des paramètres de régression et de variance des équations (3.10) et (3.11); $f_{Y_i|b_i}$ est la densité de probabilité multivariée gaussienne du marqueur conditionnelle aux effets aléatoires de moyenne $X_{ij}\beta + Z_{ij}b_i$ et de variance $\sigma_e^2 I_{n_i}$, et f_{b_i} est la densité de probabilité multivariée gaussienne de moyenne 0 et de matrice de variance covariance B . Ainsi la vraisemblance est :

$$L(\theta) = \sum_{i=1}^N \log \int_{R^{n_b}} f_{Y_i|b_i}(Y_i|b; \theta) h_i(T_i|b; \theta)^{\delta_i} S_i(T_i|b; \theta) f_{b_i}(b; \theta) db, \quad (3.12)$$

$$\text{avec } f_{Y_i|b_i}(Y_i|b; \theta) = \prod_{j=1}^{n_i} f_{Y_{ij}|b_i}(Y_{ij}|b; \theta)$$

$$\text{et } S_i(T_i|b; \theta) = \exp \left(- \int_0^{T_i} h_i(s|b; \theta) ds \right)$$

où S_i est la fonction de survie. La log-vraisemblance peut être maximisée par un algorithme hybride combinant l'algorithme EM et celui de quasi-Newton utilisant pour le calcul des intégrales sur les effets aléatoires ou de leurs dérivées la quadrature pseudo-adaptative de Gauss-Hermite [Rizopoulos, 2012]. L'ensemble de cette estimation est implémenté dans le package R *JM*.

Adéquation du modèle

L'évaluation de l'adéquation aux données du modèle conjoint à effet aléatoire partagé se fait à la fois sur le processus longitudinal et le processus de survie.

Processus longitudinal : Parce que le modèle conjoint à effet aléatoire partagé suppose que la sortie d'étude consécutive à la survenue de l'évènement dépend des effets aléatoires, il est préférable de comparer les prédictions moyennes $\hat{Y}_{ij}$ conditionnelles aux effets aléatoires aux observations [Commenges and Jacqmin-Gadda, 2016].

$$\hat{Y}_{ij} = \tilde{E}(Y_i(t)|X_i(t), b_i)$$

Processus de survie : L'adéquation est évaluée de plusieurs manières :

Qualité d'ajustement du modèle par la comparaison entre les courbes de survie prédites $\hat{S}_i$ par rapport aux estimations de Kaplan-Meier ;

$$\text{où } \hat{S}_i = \frac{1}{N} \sum_{i=1}^N S_i(t|\hat{b}_i, W_i; \hat{\theta})$$

Qualité globale du modèle par la comparaison graphique d'une distribution exponentielle de paramètre 1 à celle des résidus de Cox-Snell A_i .

$$\text{où } A_i \left(\tilde{T}_i|\hat{b}_i, W_i; \hat{\theta} \right) = \int_0^{\tilde{T}_i} h_i(t|\hat{b}_i, W_i; \hat{\theta}) dt$$

Hypothèse de log-linéarité en $\kappa(\cdot)$ par l'évaluation de la structure de dépendance entre Y et T par un graphique des résidus de martingale $\delta_i - A_i \left(\tilde{T}_i|\hat{b}_i, W_i; \hat{\theta} \right)$ versus la valeur prédite de la fonction $\kappa(\cdot)$.

Enfin, la comparaison de plusieurs modèles conjoints avec différentes fonctions de lien ($\kappa(\cdot)$) entre Y et T [Commenges and Jacqmin-Gadda, 2016].

3.2.2 Modèle conjoint à classes latentes

Le modèle conjoint à classes latentes a été décrit pour la première fois par Lin [Lin et al., 2002]. Ce modèle conjoint suppose que la population d'étude est hétérogène, composée de G sous-groupes de populations homogènes non observées représentées par G classes latentes au sein desquelles les patients partagent à la fois la même évolution du marqueur et le même risque d'évènement. Le modèle repose sur l'hypothèse d'indépendance conditionnelle selon laquelle toute la corrélation entre le marqueur Y et le risque d'évènement T est entièrement capturée par les classes latentes.

Spécification du modèle

Le modèle conjoint à classes latentes combine trois sous-modèles :

un modèle logistique multinomiale pour la probabilité d'appartenance aux classes comme définit dans l'équation (3.5) :

$$\pi_{ig} = P(c_i = g|X_{1i}) = \frac{\exp(\zeta_{0g} + X_{1i}^T \zeta_{1g})}{\sum_{l=1}^G \exp(\zeta_{0l} + X_{1i}^T \zeta_{1l})}$$

avec $\zeta_{0G} = 0$ et $\zeta_{1G} = 0$ pour assurer l'identifiabilité du modèle logistique,

un modèle linéaire mixte pour l'évolution du marqueur Y spécifique à chaque classe g :

$$Y_i(t_{ij}) = X_{2ij}^T \beta_g + Z_{ij}^T b_{ig} + \varepsilon_{ij} \quad (3.13)$$

avec $b_{ig} \sim N(0, B_g)$ et $\varepsilon_{ij} \sim N(0, \sigma_e^2)$, et

un modèle de survie pour le risque d'évènement spécifique à chaque classe g :

$$h_i(t|c_i = g; \alpha_g) = h_0(t; \zeta_g) \exp(W_i^T \alpha_g) \quad (3.14)$$

avec une fonction du risque de base $h_0(\cdot)$ spécifique ou non aux classes et α_g le vecteur des paramètres de régression spécifique aux classes associé aux variables W_i .

Estimation du modèle

Selon l'hypothèse d'indépendance conditionnelle, la distribution conjointe peut s'écrire :

$$f_{Y_i, T_i, \delta_i}(Y_i, T_i, \delta_i; \theta_G) = \sum_{g=1}^G f_{Y_i|c_i}(Y_i|c_i = g; \theta_G) P(T_i, \delta_i|c_i = g; \theta_G) \pi_{ig}$$

où $Y_i = (Y_{i1}, \dots, Y_{in_i})^T$ est le vecteur transposé des observations du marqueur Y_i pour le patient i à la mesure j , $f_{Y_i|c_i}(Y_i|c_i = g)$ est la densité des données répétées du marqueur dans la classe g qui suit une loi normale multivariée de moyenne $X_{2i}\beta_g$ et de matrice de variance $V_{ig} = \sigma_g Z_i B Z_i^T + \sigma_\epsilon^2 I_{n_i}$. Le vecteur θ_G contient l'ensemble des paramètres des équations (3.5), (3.13) et (3.14). Ainsi la vraisemblance est de la forme :

$$L(\theta_G) = \prod_{i=1}^N \sum_{g=1}^G f_{Y_i|c_i}(Y_i|c_i = g; \theta_G) h_i(T_i|c_i = g; \theta_G)^{\delta_i} S_i(T_i|c_i = g; \theta_G) \pi_{ig} \quad (3.15)$$

où π_{ig} est la probabilité d'appartenir à la classe g définie en (3.13), $h_i(t|c_i = g)$ la fonction de risque instantanée définie en (3.14) et $S_i(T_i|c_i = g)$ la fonction de survie spécifique à la classe g . Pour un nombre de classes latentes fixé, la vraisemblance peut être maximisée par l'algorithme de Marquardt modifié décrit précédemment [Commenges and Jacqmin-Gadda, 2016].

Discrimination du modèle

Comme pour les modèles mixtes à classes latentes, une classification *a posteriori* pour chaque patient d'appartenir à une classe latente peut être dérivé des modèles conjoints à classes latentes. Cependant, cette probabilité *a posteriori* peut être :

-Soit conditionnelle à la fois aux données longitudinales et de survie, utile pour évaluer l'adéquation du modèle :

$$\hat{\pi}_{ig}^{Y,T} = P(c_i = g|Y_i, T_i, \delta_i; \hat{\theta}_G) = \frac{\hat{\pi}_{ig} f_{Y_i|c_i}(Y_i|c_i = g; \hat{\theta}_G) h_i(T_i|c_i = g; \hat{\theta}_G)^{\delta_i} S_i(T_i|c_i = g; \hat{\theta}_G)}{\sum_{m=1}^G \hat{\pi}_{im} f_{Y_i|c_i}(Y_i|c_i = m; \hat{\theta}_G) h_i(T_i|c_i = m; \hat{\theta}_G)^{\delta_i} S_i(T_i|c_i = m; \hat{\theta}_G)} \quad (3.16)$$

-Soit conditionnelle uniquement aux données longitudinales, utile pour faire de la prédiction à partir des mesures répétées du marqueur [Commenges and Jacqmin-Gadda, 2016] :

$$\hat{\pi}_{ig}^Y = P(c_i = g|Y_i, \delta_i; \hat{\theta}_G) = \frac{\hat{\pi}_{ig} f_{Y_i|c_i}(Y_i|c_i = g; \hat{\theta}_G)}{\sum_{m=1}^G \hat{\pi}_{im} f_{Y_i|c_i}(Y_i|c_i = m; \hat{\theta}_G)} \quad (3.17)$$

Ainsi, à partir de ces probabilités *a posteriori*, chaque patient est affecté dans la classe à laquelle il a la plus grande probabilité d'appartenir : $\tilde{c}_i^{Y,T} = \argmax_g(\hat{\pi}_{ig}^{Y,T})$ et $\tilde{c}_i^Y = \argmax_g(\hat{\pi}_{ig}^Y)$.

Adéquation du modèle

L'adéquation du modèle conjoint à classe latente concerne à la fois les prédictions longitudinales et de survie. Pour la partie longitudinale, l'évaluation de l'adéquation est faite de la même façon que pour les modèles mixtes à classes latentes (voire section 3.2). Quand à la partie survie, on peut comparer les courbes de survie prédites obtenues à partir des fonctions de survies conditionnelles aux classes $\hat{S}_{ig}(t)$ aux courbes de survie estimées par Kaplan-Meier :

$$\hat{S}_{ig}(t) = \sum_{i=1}^N \hat{\pi}_{ig} S_i(t|c_i = g; \hat{\theta}_G)$$

où $S_i(t|c_i = g; \hat{\theta}_G)$ dérive de l'équation (3.14).

Comme pour les prédictions sur Y , les $\hat{S}_{ig}(t)$ peuvent aussi être moyennées par patient ou par classe.

Hypothèse d'indépendance conditionnelle

Le modèle conjoint à classe latente repose sur l'hypothèse centrale que la corrélation entre le marqueur longitudinal et le temps d'évènement est capturée par la structure des classes latentes [Jacqmin-Gadda et al., 2010]. Cette hypothèse est évaluée par le test du score dont le principe consiste à évaluer l'absence de dépendance résiduelle entre le marqueur longitudinal et le temps d'évènement, conditionnellement aux classes latentes. Sous l'hypothèse alternative H_1 , le sous-modèle de survie de l'équation (3.14) est modifié par l'inclusion des effets aléatoires partagés b_i en plus des classes latentes :

$$h_i(t|c_i = g; \alpha_g) = h_0(t; \zeta_g) \exp(W_i^T \alpha_g + b_{ig} \eta) \quad (3.18)$$

où η est le vecteur de paramètre associant les effets aléatoires du modèle longitudinal (3.14) au temps d'évènement.

Ainsi, l'évaluation de l'hypothèse d'indépendance conditionnelle consiste à tester $H_0 : \eta = 0$ (*versus* $H_1 : \eta \neq 0$), et la statistique du score test correspondante est définie :

$$U(0, \theta_G) = \sum_{i=1}^N \sum_{g=1}^G \hat{\pi}_{ig}^{Y,T} (\delta_i - A_{ig}(T_i; \theta_G)) E(b_{ig}|c_i = g, Y_i; \theta_G) \quad (3.19)$$

où $A_{ig}(T_i; \theta_G)$ est la fonction de risque cumulée en T_i et $E(b_{ig}|c_i = g, Y_i; \theta_G)$ est l'espérance *a posteriori* estimée à partir de l'estimateur empirique bayésien $\hat{b}_{ig}$. La statistique du test du score est la dérivée par rapport à η de la log-vraisemblance calculée en η du modèle (3.19). Elle correspond à une estimation de la covariance entre les résidus du modèle de survie et les estimateurs bayésiens empiriques des effets aléatoires pondérée par les probabilités *a posteriori* d'appartenance aux classes [Commenges and Jacqmin-Gadda, 2016, Jacqmin-Gadda et al., 2010].

Sous $H_0 : \eta = 0$, la statistique du score test $U(0, \hat{\theta}_G)^T \text{var}(U)^{-1} U(0, \hat{\theta}_G)$ suit asymptotiquement une distribution du χ^2 à m degrés de liberté où m est la dimension du vecteur η .

Sélection du modèle

Il est recommandé de sélectionner le nombre optimal de classes latentes suivant les objectifs de l'étude sur un faisceau d'éléments, qui en plus du BIC , doit tenir compte de la discrimination des classes (entropie), du $ICL - BIC$, de l'adéquation correcte du modèle, de la validation de l'hypothèse d'indépendance conditionnelle aux classes et/ou de la pertinence clinique des classes [Commenges and Jacqmin-Gadda, 2016].

4 Analyse de trajectoire : Impact de la perte de poids sur la progression de l'ataxie

4.1 Résumé

4.1.1 Rationnel

De nombreuses études incluant des patients atteints d'ataxie mais aussi des modèles de souris transgéniques sur les SCA suggèrent que dans les pathologies à expansion de polyglytamine, la progression de la maladie serait précédée par une perte de poids [Saute et al., 2012, Rüb et al., 2006, Jafar-Nejad et al., 2011, Hübener et al., 2012, Boy et al., 2009, Mähler et al., 2014]. Cependant, la majorité de ces études était transversales ou rétrospectives avec un nombre limité de patients [Saute et al., 2012]. Dans ce travail, nous présentons la première étude longitudinale sur l'évolution du poids chez des patients SCA et l'identification de sous-groupes d'évolution différente du poids.

4.1.2 Méthodes

Parmi les 525 patients SCA inclus dans la cohorte EUROSCA, nous avons fait une analyse poolée des 384 patients (80 SCA1, 125 SCA2, 105 SCA3 et 74 SCA6) ayant au moins 3 mesures du poids. L'indice de masse corporelle calculé pour chaque patient en divisant le poids sur la taille au carré ($IMC = \frac{poids}{taille^2}$) a été utilisé comme critère principal. Le score SARA utilisé pour mesurer la sévérité de l'ataxie a été considéré comme critère secondaire.

Evolution de l'indice de masse corporelle (IMC)

Pour étudier l'évolution de l'IMC en fonction de l'âge, nous avons utilisé le modèle mixte [Laird and Ware, 1982] décrit dans le chapitre précédent. Afin que l'intercept du modèle puisse être interprété et éviter les problèmes numériques liés à l'estimation des paramètres (de variance notamment), nous avons centré l'âge à 18 ans (minimum dans la cohorte) et l'avons considéré en dizaine d'année : $age = (\frac{Age-18}{10})$. En supposant une évolution cubique avec l'âge et en incluant des effets aléatoires sur le niveau initial (b_{0i}) et les fonctions du temps (âge (b_{1i}) et âge au carré (b_{2i})), et un processus autorégressif d'ordre 1 pour tenir compte de la variabilité inter et intra-patients, le modèle mixte s'écrit pour le patient i à la mesure répétée j :

$$IMC_{ij} = (\beta_0 + b_{0i}) + (\beta_1 + b_{1i}) \times age_{ij} + (\beta_2 + b_{2i}) \times age_{ij}^2 + \beta_3 \times age_{ij}^3 + \varepsilon_{ij} \quad (4.1)$$

avec

$$b_i = (b_{0i}, b_{1i}, b_{2i})^T \sim N \left(0, B = \begin{bmatrix} \sigma_{b_0}^2 & \sigma_{b_{01}} & \sigma_{b_{02}} \\ & \sigma_{b_1}^2 & \sigma_{b_{12}} \\ & & \sigma_{b_2}^2 \end{bmatrix} \right)$$

et $\varepsilon_{ij} \sim N(0, \Sigma_{ij})$

où les effets aléatoires b_i sont corrélés entre eux, de variance non structurée B ; ε_{ij} sont les erreurs résiduelles corrélées entre elles, et sont la somme d'un processus gaussien autorégressif capturant la corrélation résiduelle entre les mesures successives de l'IMC ($w_i(\text{age}_{ij})$) et d'une erreur de mesure indépendante (ϵ_{ij}) liée aux instruments de mesure du poids et de la taille :

$$\varepsilon_{ij} = w_i(\text{age}_{ij}) + \epsilon_{ij} \text{ avec } \epsilon_{ij} \sim N(0, \sigma_\epsilon^2) \text{ et } w_i \sim N(0, \sigma_w^2)$$

où σ_ϵ^2 est la variance de l'erreur de mesure et σ_w^2 celle du processus autorégressif avec une corrélation entre deux mesures ($j \neq k$) successives pour un patient i de :

$$\rho = [w_i(\text{age}_{ij}), w_i(\text{age}_{ik})] = e^{(-\varrho|\text{age}_{ij}-\text{age}_{ik}|)}$$

avec $\varrho > 0$

Les effets aléatoires b_i , l'erreur autorégressive $w_i(\text{age}_{ij})$ et l'erreur ϵ_{ij} sont indépendantes. Nous avons utilisé la fonction *hlme* du package R *lcm* pour estimer le modèle mixte [Proust-Lima et al., 2017].

Sous-groupe (ou trajectoire) d'évolution de l'IMC

Pour identifier des sous-groupes d'évolution différente de l'IMC, nous avons utilisé un modèle mixte à classe latente [Muthén and Shedden, 1999, Commenges and Jacqmin-Gadda, 2016] décrit dans la section 3.1.2. Pour cette analyse, nous n'avons inclut aucun prédicteur dans la probabilité d'appartenir aux classes latentes, si bien que

$$\pi_{ig} = P(c_i = g) = \frac{e^{\zeta_{0g}}}{\sum_{l=1}^G e^{\zeta_{0l}}} \quad (4.2)$$

avec $\zeta_{0G} = 0$, G étant la classe de référence.

L'évolution de l'IMC conditionnelle à la classe latente g pour le patient i à la mesure j est définie par :

$$[IMC_{ij}|c_i = g] = (\beta_{0g} + b_{0i}) + (\beta_{1g} + b_{1i}) \times \text{age}_{ij} + (\beta_{2g} + b_{2i}) \times \text{age}_{ij}^2 + \beta_{3g} \times \text{age}_{ij}^3 + \varepsilon_{ij} \quad (4.3)$$

avec $b_i = (b_{0i}, b_{1i}, b_{2i})^T \sim N(0, B_g)$ et $\varepsilon_{ij} \sim N(0, \Sigma_{ij})$

La matrice de variance des erreurs ε_{ij} est la même que celle décrite dans l'équation (4.1) ; pour la matrice de covariance des effets aléatoires, nous avons supposé qu'elle était non structurée de variance commune entre les classes latentes : $B_g = B, \forall g \in [1, G]$:

La sélection du nombre de classes optimales a été faite par le Bayesian Information Criterion *BIC*. Pour chaque patient i en fonction de ses observations IMC_{ij} une probabilité *a posteriori* d'appartenir à la classe latente g suivant l'équation 3.9 a été calculée. Cette classification *a*

posteriori a été utilisée pour évaluer la qualité de discrimination et l'adéquation du modèle. Nous avons ensuite comparé la distribution de ces classes latentes selon les caractéristiques de la population à l'inclusion (voire tableau 1 de l'article). En fonction de la nature des variables à l'inclusion, nous avons utilisé soit un test de Chi-deux pour les variables catégorielles soit une analyse de variance pour les variables continues avec le test de Honestly Significant Difference (HSD) [Bernhardson, 1975] pour tenir compte des comparaisons multiples.

Impact des trajectoires de l'IMC sur la progression de l'ataxie

Pour évaluer si la progression de l'ataxie mesurée par le score SARA est directement liée aux différentes évolutions de l'IMC, nous avons, *a posteriori* utilisé un modèle linéaire mixte à intercept et pente aléatoires en introduisant dans celui-ci l'interaction entre le temps (*age*) et la classe latente g .

$$SARA_{ij} = X_{ij}^T \beta + b_{0i} + b_{1i} \times age_{ij} + \epsilon_{ij} \quad (4.4)$$

avec

$$b_i = (b_{0i}, b_{1i})^T \sim N \left(0, B = \begin{bmatrix} \sigma_{b_0}^2 & \sigma_{b_{01}} \\ \sigma_{b_{01}} & \sigma_{b_1}^2 \end{bmatrix} \right)$$

et $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$

où X_{ij}^T est le vecteur de variable correspondant à la classe latente d'évolution de l'IMC associé au vecteur de coefficients β des effets fixes incluant l'intercept, la pente moyenne et l'interaction entre la classe latente et le temps (*age*).

Analyse de sensibilité

En raison de l'extension de la cohorte EUROSCA après les 3 ans de suivi, il a été montré que les patients suivis au maximum 3 ans présentaient plus de taux de sortie d'étude et une progression plus rapide de la maladie que les patients suivis plus longtemps [Jacobi et al., 2015]. Dans ces circonstances, le mécanisme de sortie d'étude est dit manquant non aléatoire ou missing not at random (MNAR) car la probabilité de sortie d'étude dépend des valeurs non-observées du poids au temps t ; le mécanisme est donc informatif. Or, le modèle linéaire mixte comme le modèle mixte à classe latente supposent que les données sont manquantes au hasard. Pour s'assurer que les estimations ne sont pas biaisées, nous avons réalisé une analyse de sensibilité sous l'hypothèse de données manquantes non aléatoires en utilisant un modèle conjoint à effets aléatoires partagés [Rizopoulos, 2012] décrit dans la section 3.2.1. Ce modèle conjoint correspond aux modèles de sélection dépendant des effets aléatoires dans la terminologie utilisée pour les données longitudinales incomplètes. Nous avons supposé que le suivi a été interrompu par la sortie d'étude (dropout ou décès) entraînant des données manquantes (du poids) de façon monotone. Dans ce modèle conjoint, le modèle longitudinal était un modèle mixte (équation 4.1) et le modèle de survie, un modèle paramétrique avec une fonction du risque de base de type *Weibull* où l'évènement noté δ_i était la sortie d'étude (mesure du poids manquante ou décès), et le délai de sortie noté T_i celui entre l'inclusion et la sortie toute cause. Le modèle de survie a ainsi été défini par :

$$h_i(t|b_i, \alpha^{(a)}) = h_0(t) \exp(\kappa(b_i, t_{ij})^T \alpha^{(a)}) \quad (4.5)$$

où $h_i(t)$ est la fonction du risque de sortie d'étude, $h_0(t)$ la fonction paramétrique du risque de base *Weibull* avec $h_0(t) = \gamma \lambda^\gamma t^{\gamma-1}$ avec γ la forme du risque de base, λ l'échelle (ou scale). La

fonction $\kappa_i(b_i, t)$ définit la structure de dépendance (ou fonction de lien) entre le délai de sortie d'étude et l'évolution de l'IMC et $\alpha^{(a)}$ quantifie la mesure de l'association entre les deux processus. Si $\alpha^{(a)} = 0$, le mécanisme de sortie est MCAR [Molenberghs et al., 2015, Rizopoulos, 2012]. Nous avons supposé que le risque instantané de sortie d'étude en t dépend de la valeur courante de l'IMC en t débarrassée de l'erreur de mesure :

$$\kappa_i(b_i, t_{ij}) = \tilde{Y}_i(t_{ij}) \text{ où } \tilde{Y}_i(t_{ij}) \text{ correspond à la mesure l'IMC au temps } j \text{ pour le sujet } i.$$

Nous avons estimé les paramètres du modèle conjoint en utilisant le package R *JM*.

4.1.3 Résultat

Les résultats de ce travail ont permis de montrer que l'IMC à l'inclusion était significativement plus bas pour les patients SCA3 comparé aux autres. Nous avons également montré, que l'IMC diminuait globalement de façon cubique au cours du temps, et que cette diminution n'était pas significativement différente entre les génotypes. Pour la première fois, il a été réalisé une analyse en sous-groupe d'évolution de l'IMC qui a permis l'identification de trois sous-groupes : stable, augmentation et diminution. Le premier sous-groupe qui comprenait 59% des patients était caractérisé par une évolution plus lente et stable de l'IMC, le second (18%) correspondant aux patients qui augmentaient leur IMC au cours du temps, et enfin le dernier (23%) caractérisé par une diminution au cours du temps de l'IMC.

Selon la classification *a posteriori*, les patients qui baissaient leur IMC présentaient une atteinte clinique significativement plus sévère tant sur les symptômes cérébelleux (score SARA plus élevé) que sur les symptômes non ataxiques (nombre de signes non ataxiques plus élevé), et une forme génétique également plus sévère (expansion du nombre de répétitions du CAG plus longue) comparés aux autres sous-groupes. Enfin, nous avons montré que les patients qui baissaient leur IMC progressaient plus rapidement. Les résultats de l'analyse de sensibilité ont montré que l'estimation des paramètres obtenus à partir du modèle conjoint à effets aléatoires partagés était similaire à celles des méthodes standards (modèle mixte et modèle à classe latente), suggérant une absence de biais dans l'estimation des paramètres.

En conclusion, ce travail a permis d'identifier trois sous-groupes d'évolution différente de l'IMC et ont fourni des données solides en faveur d'une association significative entre la perte de poids et la progression de la maladie. Ces résultats suggèrent que la surveillance du poids corporel pourrait être utilisée comme un biomarqueur potentiellement utile de la progression de la maladie. En outre, les résultats fournissent plus d'informations pour aider à la sélection et la stratification des patients pour la conception des essais cliniques interventionnels.

L'ensemble de ce travail a fait l'objet d'une publication dans la revue *Movement Disorders Clinical Practice*.

4.2 Article 1 : The Movement Disorders Clinical Practice

Body Mass Index Decline Is Related to Spinocerebellar Ataxia Disease Progression

Alhassane Diallo, MD,¹ Heike Jacobi, MD,^{2,3} Tanja Schmitz-Hübsch, MD,⁴ Arron Cook, MD,⁵  Robyn Labrum, PhD,⁶ Alexandra Durr, MD, PhD,^{7,8,9} Alexis Brice, MD,^{7,8,9} Perrine Charles, MD,¹⁰ Cecilia Marelli, MD,¹¹ Caterina Mariotti, MD,¹² Lorenzo Nanetti, MD,¹² Marta Panzeri, MD,¹² Maria Rakowicz, MD, PhD,¹³ Anna Sobanska, MD,¹³ Anna Sulek, PhD,¹⁴ Ludger Schöls, MD,¹⁵ Holger Hengel, MD,¹⁵ Bela Melegh, PhD,¹⁶ Alessandro Filla, MD,¹⁷ Antonella Antenora, MD,¹⁷ Jon Infante, MD,¹⁸ José Berciano, MD,¹⁸ Bart P. van de Warrenburg, PhD,¹⁹ Dagmar Timmann, MD,²⁰ Sylvia Boesch, MD,²¹ Massimo Pandolfo, MD,²² Jörg B. Schulz, MD,^{23,24} Peter Bauer, MD,²⁵ Paola Giunti, PhD,⁵ Laszlo Baliko, MD,²⁶ Michael H. Parkinson, MBBS,⁵ Jun-Suk Kang, MD,²⁷ Thomas Klockgether, MD,²⁸ Sophie Tezenas du Montcel, MD, PhD^{1,29,*}

Abstract: Background: Spinocerebellar ataxias (SCAs) are dominantly inherited, progressive ataxia disorders. Disease progression could be preceded by weight loss.

Objectives: We aimed to study the course of weight loss in patients who had the most common SCAs (SCA1, SCA2, SCA3, and SCA6). Additional objectives were to identify subgroups of weight evolution, to determine the factors influencing these evolutions, and to assess the impact of these evolutions on disease progression.

Methods: In total, 384 patients from the EUROSCA prospective cohort study were analyzed who had SCA1, SCA2, SCA3, or SCA6 and at least 3 measurements of weight. Age was used as a time scale. Clinical outcomes were body mass index (BMI) and the Scale for the Assessment and Rating Ataxia (SARA), with scores ranging from 0 to 40. We used a linear mixed model to analyze the course of BMI and a latent class mixed model to identify subgroup BMI evolution.

Results: Overall, BMI declined over time (-0.11 ± 0.03 kg/m² per decade; $P = 0.0009$). Three subgroups of BMI evolution were identified: “decreasing BMI” ($n = 88$; 23%), “increasing BMI” ($n = 70$; 18%) and “stable BMI” ($n = 226$; 59%). Patients in the decreasing BMI group were more severely affected at baseline with higher SARA scores and a

¹Paris-Sorbonne Universités, Université Pierre et Marie Curie UMR 1136, Institut National de la Santé et de la Recherche Médicale (INSERM), Institut Pierre Louis d'Epidémiologie et de Santé Publique, Paris, France; ²Department of Neurology, University Hospital of Bonn, Bonn, Germany; ³Department of Neurology, University Hospital of Heidelberg, Heidelberg, Germany; ⁴Clinical Neuroimmunology Groupe, NeuroCure Clinical Research Center, Charité Universitätsmedizin Berlin, Berlin, Germany; ⁵Department of Molecular Neuroscience, Institute of Neurology, University College London, London, United Kingdom; ⁶Neurogenetic Laboratory, National Hospital of Neurology and Neurosurgery, University College London Hospitals, London, United Kingdom; ⁷Paris-Sorbonne Universités, Université Pierre et Marie Curie UMR S1127, INSERM, Paris, France; ⁸Centre National de la Recherche Scientifique, Institut du Cerveau et de la Moelle Épinrière UMR 7225, Assistance Publique-Hôpitaux de Paris, Paris, France; ⁹Genetic Department, Pitié-Salpêtrière University Hospital, Paris, France; ¹⁰Genetic Department, Pitié-Salpêtrière University Hospital, Assistance Publique-Hôpitaux de Paris, Paris, France; ¹¹Service de Neurologie, Centre Mémoire de Ressources et de Recherche, Hôpital Gui de Chauliac, Montpellier, France; ¹²SOSD Genetics of Neurodegenerative and Metabolic Diseases, Fondazione Istituto di Ricovero e Cura a Carattere Scientifico, Milan, Italy; ¹³Department of Clinical Neurophysiology, Institute of Psychiatry and Neurology, Warsaw, Poland; ¹⁴Department of Genetics, Institute of Psychiatry and Neurology, Warsaw, Poland; ¹⁵Department of Neurodegeneration and Hertie Institute for Clinical Brain Research, University of Tübingen and Deutsches Zentrum für Neurodegenerative Erkrankungen, Tübingen, Germany; ¹⁶Department of Medical Genetics and Szentagothai Research Center, University of Pécs, Pécs, Hungary; ¹⁷Department of Neuroscience and Reproductive and Odontostomatological Sciences, Federico II University, Naples, Italy; ¹⁸Neurology Service, University Hospital Marqués de Valdecilla, University of Cantabria and Centro de Investigación Biomédica en Red de Enfermedades Neurodegenerativas (CIBERNED), Santander, Spain; ¹⁹Department of Neurology, Donders Institute for Brain, Cognition and Behavior, Radboud University Medical Center, Nijmegen, the Netherlands; ²⁰Department of Neurology, University Clinic Essen, University of Duisburg-Essen, Duisburg, Germany; ²¹Department of Neurology, Medical University, Innsbruck, Innsbruck, Austria; ²²Neurology Service, Laboratory of Experimental Neurology, Université Libre de Bruxelles Hôpital Erasme, Brussels, Belgium; ²³Department of Neurology, Rheinisch-Westfälische Technische Hochschule Aachen University, Aachen, Germany; ²⁴Jülich Aachen Research Alliance-Translational Brain Medicine, Aachen, Germany; ²⁵Institute of Medical Genetics and Applied Genomics, University of Tübingen, Tübingen, Germany; ²⁶Department of Neurology, Zala County Hospital, Zalaegerszeg, Hungary; ²⁷Department of Neurology, University of Frankfurt, Frankfurt am Main, Germany; ²⁸Department of Neurology and German Center for Neurodegenerative Diseases, University Hospital of Bonn, Bonn, Germany; ²⁹Biostatistics Unit, Groupe Hospitalier Pitié-Salpêtrière, AP-HP, Paris, France

***Correspondence to:** Dr. Sophie Tezenas du Montcel, Sorbonne Universités, Université Pierre et Marie Curie (UPMC) Univ Paris 06, UMR S1136, INSERM U 1136, Institut Pierre Louis d'Epidémiologie et de Santé Publique, F-75013, Paris, France; and AP-HP Biostatistics Unit, Groupe Hospitalier Pitié-Salpêtrière, F-75013, Paris, France; E-mail: sophie.tezenas@aphp.fr

Keywords: body mass index, latent class mixed model, longitudinal data, spinocerebellar ataxia.

Relevant disclosures and conflicts of interest are listed at the end of this article.

Received 17 February 2017; revised 1 June 2017; accepted 17 June 2017.

Published online xx XXXXX 2017 in Wiley InterScience (www.interscience.wiley.com). DOI:10.1002/mdc3.12522

higher frequency of non-ataxia signs (especially motor symptoms) compared with those in the other groups. Weight loss was associated with faster disease progression (5.7 ± 0.7 SARA points per decade; $P = 0.036$).
Conclusions: The current data have substantial implications for the design of future interventional studies in SCA, as they provide a basis for patient stratification and emphasize the usefulness of BMI as a biomarker for monitoring disease progression.

Spinocerebellar ataxias (SCAs) are a clinically and genetically heterogeneous group of autosomal-dominant, inherited, progressive ataxia disorders. Several genetically different SCAs have been defined, and the most common are SCA1, SCA2, SCA3, and SCA6.¹ These ataxias are caused by translated CAG repeat expansions, which code for elongated polyglutamine tracts within the different proteins associated with each type. Data from a large prospective study of these disorders yielded important insights into their natural history. Progressive ataxia is the leading symptom in these disorders and increases linearly over time, with different rates in each genotype. Progression is faster in SCA1, intermediate in SCA2 and SCA3, and slower in SCA6. Factors associated with faster disease progression have been identified in all SCAs except SCA3. These factors include older age at baseline and longer repeat expansions in SCA1, younger age at onset in SCA2, and lower baseline Scale for the Assessment and Rating Ataxia (SARA) scores in both SCA2 and SCA6.²

The occurrence of gait ataxia is preceded by a preclinical stage, resulting in dysfunction, degeneration, and metabolic changes in several nervous system compartments.^{3,4} Degenerative processes affecting brain structures other than the cerebellum may explain extracerebellar signs in SCAs,⁵ such as dysphagia and weight loss.^{6,7}

A significant decline in body weight directly linked with increased CAG repeat length has been observed in Huntington's disease (HD).⁸ Moreover, several studies that included data from patients with SCA^{9,10} and transgenic mouse models of SCA^{11–14} have reported the presence of weight loss in SCAs. Dysphagia¹⁰ and an increased catabolic state¹⁵ have been reported as mechanisms underlying weight loss in SCAs.

The previous studies were either cross-sectional or retrospective, had case-control designs, and included small patient numbers. Here, we explore data from a longitudinal EUROSCA cohort with a long-term follow-up on body weight and height in relation to ataxia progression.

The objectives of the current analysis were to: (1) specify the course of weight loss, (2) identify different subgroups of weight evolution, (3) determine the factors associated with weight loss, and (4) quantify the association between disease progression and weight loss.

Patients and Methods

Standard Protocol Approvals, Registration, and Patients Consent

The ethics committees of the participating centers approved the study (clinicaltrials.gov identifier NCT02440763). At

enrolment, a written consent form was obtained from all study participants.

Study Population

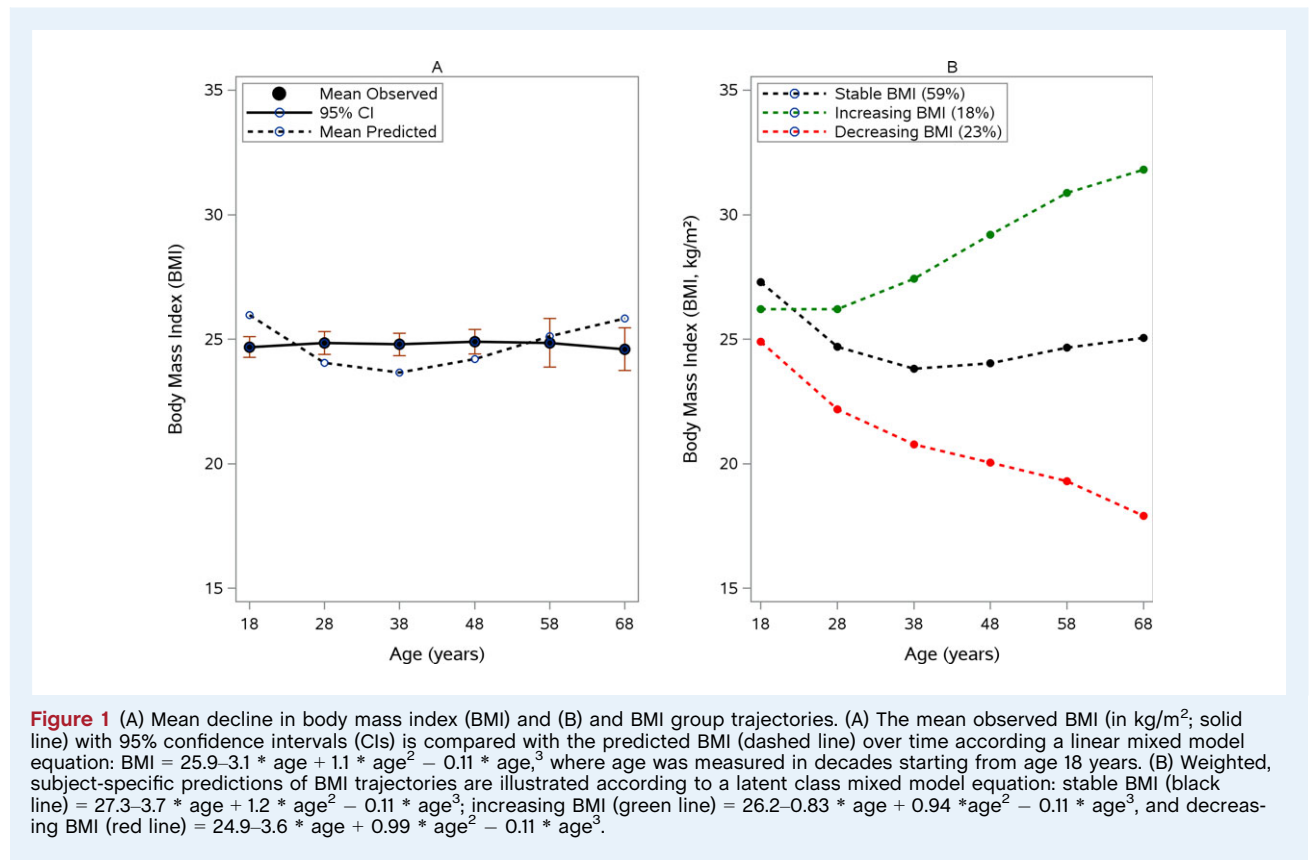
Of 525 patients with SCA who were enrolled between July 1, 2005, and Aug 31, 2006, in the multicenter (17 European centers) EUROSCA prospective cohort study,¹⁶ a subsample of 384 patients who had at least 3 measures of weight over a maximum of 6 years of follow-up were analyzed (Fig. S1). These patients were diagnosed with progressive, otherwise unexplained ataxia and had a positive molecular genetic test for SCA1 ($n = 80$), SCA2 ($n = 125$), SCA3 ($n = 105$), or SCA6 ($n = 74$).

Clinical and Genetics Evaluations

Demographic data included age, sex, age at onset, and disease duration. The following variables were measured at baseline: weight and height; disease stage; symptoms other than ataxia according to Inventory of Non-ataxia Signs (INAS) scores (range, 0–16) signs, including the extended INAS score, such as dysphagia and double vision¹⁷; anxiety or depression status measured by Patient Health Questionnaire-9 (PHQ-9) self-rated items¹⁸; the use of any psychiatric medication among antipsychotics, anxiolytics, hypnotics and sedatives, antidepressants, psychostimulants, agents used for attention deficit hyperactivity disorder, and nootropics; and the SARA score. Weight and SARA score were also measured at subsequent visits at 1, 2, and 3 years (± 3 months). After the initial 3-year observation period, the study participants entered an extension phase in which study assessments were done in connection with routine visits, resulting in irregular intervals between the visits. Body mass index (BMI) was calculated with the formula ($\text{weight}/(\text{height})^2$). Repeat lengths of the expanded and normal alleles were determined at the Institute of Medical Genetics and Applied Genomics of the University of Tübingen (Tübingen, Germany).

Statistical analysis

Descriptive statistics (frequencies and percentages or means $\pm$ standard deviations) were used to describe demographics and clinical characteristics of the participants at baseline. BMI at baseline was compared between genotypes using an analysis of variance with Tukey's Honestly Significant Difference test to account for



multiple comparisons. To describe changes in BMI over time, the standard linear mixed (LM) model¹⁹ was used with the cubic function of age as a time scale and without any adjustment for covariates. The cubic function was the best of the tested models (linear, quadratic, and cubic models). Correlated, autoregressive, random intercept, linear and quadratic slope accounted for inter-patient variability. The age variable was centered at 18 years (the minimum age in the cohort) and is indicated in 10-year increments (i.e., ages 18, 28, 38, 48 years, etc.). We tested the effect of genotype on BMI evolution by interaction between the given factor and the time variable.

To identify subgroups of participants who exhibited different trajectories of BMI, a latent class mixed (LCM) model that accounted for individual and latent group structure variability through random effects was used. The same fixed and random-effects that were included in the LM model were used in the LCM model.¹⁹ We selected the best-fitting model with the optimal number of latent classes using a compromise between the Bayesian Information Criterion, a mean posterior probability >0.70 for each latent class, and a number of patients in each class >5.²⁰ Distributions of the baseline variables across these classes were compared a posteriori using a Chi-squared (χ^2) test for the categorical variables and an analysis of variance for continuous variables with the Honestly Significant Difference test to account for multiple comparisons. To assess whether the rate of change in the SARA score was directly related to these classes, a posteriori LM model was used for the SARA score, and the interaction between class variable and time was tested.

Furthermore, a previous report on long-term disease progression revealed a substantial dropout rate, mainly for disease-related reasons during the open extension period, which followed the initial 3-year period with annual visits. Patients who had a maximum follow-up of 3 years had a high dropout rate and faster disease progression.² In these patients, the dropout rate is informative, and standard estimation methods like LM models and LCM models are biased. Informative dropout leads to a “missing not at random” (MNAR) assumption that the observation probability depends on unobserved outcomes.¹⁹ To account for an informative dropout process and ensure that estimates from “missing at random” assumptions are not biased, a sensitivity analysis under an MNAR assumption was performed using joint models for longitudinal data and the time to dropout.²⁰ The estimates from the joint model under MNAR were then compared with those obtained from the LM model under a missing at random assumption. All data analyses were performed using R version 3.2.0 (R Foundation for Statistical Computing Vienna, Austria) and SAS (version 9.4; SAS Institute, Cary, NC).

Results

Baseline Characteristics

The 384 patients included in the current study had a total of 1795 visits, with a median of 3 visits (interquartile range, 3–4 visits) for each patient. Baseline characteristics of the patients are

TABLE 1 Population characteristics at baseline

Characteristic	Frequency/No. (%) or Mean $\pm$ SD			
	SCA1, n = 80	SCA2, n = 125	SCA3, n = 105	SCA6, n = 74
Men	51 (64)	59 (47)	55 (52)	40 (54)
Age, y	46 $\pm$ 12	46 $\pm$ 14	49 $\pm$ 12	64 $\pm$ 11
Age at onset, y	38 $\pm$ 11	35 $\pm$ 13	38 $\pm$ 11	54 $\pm$ 10
Disease duration, y	8 $\pm$ 5	11 $\pm$ 6	11 $\pm$ 6	10 $\pm$ 6
CAG*	47 $\pm$ 5	39 $\pm$ 3	67 $\pm$ 4	22 $\pm$ 1
Diabetes	3 (4)	5 (4)	2 (2)	5 (7)
Cancer	0 (0)	0 (0)	2 (4)	2 (6)
Gastrointestinal diseases	7 (9)	7 (6)	8 (8)	10 (10)
BMI, kg/m ²	24.9 $\pm$ 4	25.1 $\pm$ 4	23.2 $\pm$ 4	25.4 $\pm$ 4
Underweight	13 (17)	13 (11)	13 (14)	13 (14)
Normal weight	34 (46)	62 (51)	57 (62)	34 (44)
Overweight	18 (24)	34 (28)	16 (17)	18 (32)
Obese	10 (13)	12 (10)	6 (7)	10 (11)
SARA score	12.9 (7)	14.9 (7)	13.5 (7)	14.6 (6)
Disease stage				
1	56 (70.9)	79 (69.3)	49 (48.0)	34 (47.9)
2	19 (24.1)	24 (21.1)	37 (36.3)	33 (46.5)
3	4 (5.1)	11 (9.6)	16 (15.7)	4 (5.6)
Anxiety or depression, yes	37 (46.3)	59 (47.2)	49 (46.7)	35 (47.3)
Any psychotropic use, yes	15 (18.8)	18 (14.4)	20 (19.1)	7 (9.5)
INAS score, no. of signs	4.4 $\pm$ 2	4.1 $\pm$ 2	4.8 $\pm$ 2	2.0 $\pm$ 1
Hyperreflexia, yes	56 (70.0)	18 (14.5)	39 (37.9)	15 (20.3)
Areflexia, yes	13 (16.2)	78 (62.9)	59 (56.2)	16 (21.6)
Extensor plantar, yes	34 (46.6)	38 (33.3)	41 (39.4)	2 (2.8)
Spasticity, yes	49 (61.3)	13 (10.7)	43 (41.3)	9 (12.5)
Paresie, yes	15 (18.8)	18 (14.5)	23 (22.1)	5 (6.8)
Muscle atrophy, yes	17 (21.2)	25 (20.3)	38 (36.9)	9 (12.3)
Fasciculations, yes	28 (34.5)	51 (41.1)	33 (31.4)	1 (1.4)
Myoclonus, yes	1 (1.2)	18 (14.5)	4 (3.8)	0 (0.0)
Rigidity, yes	1 (1.2)	7 (5.6)	11 (10.7)	6 (8.1)
Chorea/dyskinesia, yes	3 (3.8)	11 (8.9)	7 (6.7)	2 (2.7)
Dystonia, yes	5 (6.2)	18 (14.4)	20 (19.0)	3 (4.1)
Resting tremor, yes	6 (7.5)	17 (13.8)	4 (3.8)	1 (1.4)
Sensory symptoms, yes	45 (60.0)	88 (72.1)	65 (66.3)	36 (50.0)
Urinary dysfunction, yes	25 (31.2)	50 (40.3)	49 (47.1)	26 (35.1)
Cognitive impairment, yes	12 (15.0)	32 (25.8)	19 (18.8)	7 (9.5)
Brainstem oculomotor signs, yes	29 (36.2)	44 (35.5)	51 (49.5)	11 (14.9)
Dysphagia, yes	42 (53.2)	62 (54.4)	59 (57.8)	39 (54.9)
Double vision, yes	13 (16.2)	26 (20.8)	58 (55.2)	36 (48.6)

SD, standard deviation; SCA, spinocerebellar ataxia; BMI, body mass index; SARA, Scale for the Assessment and Rating Ataxia; INAS, Inventory of Non-ataxia Signs.

*Values indicate the length of the expanded allele (repeat units).

provided in Table 1. BMI at baseline differed between SCA types ($P < 0.05$), with a significantly lower BMI in patients who had SCA3 compared with all others.

Rate of BMI Change

Based on an LM model, BMI declined cubically over time (-0.11 ± 0.03 kg/m² per decade; $P = 0.0009$) (Fig. 1A). The rate of decline did not differ significantly between the 4 SCA genotypes ($P = 0.4400$).

Trajectories of BMI and their Relationship to Ataxia Severity and Non-ataxia Signs

To identify different trajectories of BMI, the LCM model with several latent classes, ranging from 1 to 5, were estimated. The model with the optimal number of classes

selected by the compromise criterion included 3 classes (Tables S1 and S2). This model identified 3 different BMI evolutions (Fig. 1B). Class 1 ($n = 226$; 59%) was characterized by stable BMI evolution over time; Class 2 ($n = 70$; 18%) corresponded to patients who had an increase in BMI over time; and Class 3, which included 23% of patients ($n = 88$), was characterized by a decrease in BMI over time. According to the posterior classification provided in Table 2, the 3 classes differed for the most part on the baseline characteristics tested. Pairwise comparisons demonstrated that patients in the decreasing BMI group, compared with those in the increasing and stable BMI groups, were more severely affected at baseline and had higher SARA scores (16 ± 8 points), were more often permanently dependent on a wheelchair (19%), were younger at the time of disease onset (28.1 ± 9 years), had double vision (50%), and had more CAG repeats in the SCA1 (51 ± 6 CAG repeats) and SCA2 (43 ± 3 repeat units) cohorts. These patients had also a

TABLE 2 Description of the 3 posterior latent classes according to the patient's baseline characteristics

Characteristic	No. (%) or Mean $\pm$ SD*			P value†
	Decreasing BMI, n = 88	Increasing BMI, n = 70	Stable BMI, n = 226	
Genotype				<0.0001‡
SCA1	19 (21.6)	20 (28.6)	41 (18.1)	
SCA2	30 (34.1)	27 (38.6)	68 (30.1)	
SCA3	38 (43.2)	12 (17.1)	55 (24.3)	
SCA6	1 (1.1)	11 (15.7)	62 (27.4)	
Women	47 (53.4)	41 (58.6)	91 (40.3)	0.0095‡
Age, y	38.4 $\pm$ 11.8 ^a	50.1 $\pm$ 12.0 ^b	55.8 $\pm$ 12.1 ^c	< 0.0001‡
Age at onset, y	27.7 $\pm$ 9.2 ^a	40.6 $\pm$ 11.5 ^b	45.5 $\pm$ 11.5 ^c	< 0.0001‡
Disease duration, y	10.7 $\pm$ 6.3	9.5 $\pm$ 5.8	10.2 $\pm$ 5.3	0.4470
SARA score	15.9 $\pm$ 7.5 ^a	12.8 $\pm$ 7.4 ^b	13.6 $\pm$ 6.6 ^b	0.0108‡
Disease stage				0.0241‡
1	47 (55.3)	40 (59.7)	131 (61.2)	
2	22 (25.9)	22 (32.8)	69 (32.2)	
3	16 (18.8)	5 (7.5)	14 (6.5)	
Diabetes	0 (0.0)	5 (7.5)	10 (4.8)	0.0569
Cancer	1 (4.0)	0 (0.)	3 (3.8)	0.6202
Gastrointestinal diseases	7 (8.2)	9 (13.4)	16 (7.6)	0.3361
CAG§				
SCA1	51.3 $\pm$ 6.3 ^a	45.4 $\pm$ 3.8 ^b	45.7 $\pm$ 3.7 ^b	<0.0001‡
SCA2	42.5 $\pm$ 3.4 ^a	39.2 $\pm$ 2.9 ^b	37.7 $\pm$ 2.4 ^b	< 0.0001‡
SCA3	70.9 $\pm$ 3.1 ^a	68.6 $\pm$ 3.8 ^{a,b}	67.1 $\pm$ 4.2 ^b	< 0.0001‡
SCA6	22.0	22.3 $\pm$ 0.5	22.5 $\pm$ 1.2	0.4904
Anxiety or depression, yes	44 (50.0)	36 (51.4)	100 (44.2)	0.4597
Any psychotropic use, yes	14 (15.9)	14 (20.0)	32 (14.2)	0.4991
INAS score	5.2 $\pm$ 2.3 ^a	3.8 $\pm$ 2.2 ^b	3.5 $\pm$ 2.1 ^b	< 0.0001‡
Hyperreflexia, yes	41 (47.1)	19 (27.1)	69 (30.4)	0.0087‡
Areflexia, yes	35 (39.8)	29 (41.4)	102 (45.3)	0.6300
Extensor plantar, yes	34 (41.0)	26 (38.8)	55 (25.9)	0.0175‡
Spasticity, yes	38 (43.7)	17 (24.6)	59 (26.6)	0.0071‡
Paresis, yes	19 (21.6)	12 (17.1)	30 (13.5)	0.2030
Muscle atrophy, yes	34 (39.1)	12 (17.1)	43 (19.4)	<0.0001‡
Fasciculations, yes	37 (42.5)	24 (34.3)	52 (23.1)	0.0022‡
Myoclonus, yes	10 (11.4)	4 (5.7)	9 (4.0)	0.0488‡
Rigidity, yes	7 (8.0)	3 (4.3)	15 (6.7)	0.6342
Chorea/dyskinesia, yes	11 (12.5)	2 (2.9)	10 (4.4)	0.0124‡
Dystonia, yes	20 (22.7)	4 (5.7)	22 (9.7)	0.0013‡
Resting tremor, yes	11 (12.8)	5 (7.1)	12 (5.4)	0.0806
Sensory symptoms, yes	50 (61.7)	49 (72.1)	135 (61.9)	0.2881
Urinary dysfunction, yes	29 (33.3)	33 (47.1)	88 (39.1)	0.2115
Cognitive impairment, yes	18 (20.7)	14 (20.0)	38 (17.0)	0.6937
Brainstem oculomotor signs, yes	51 (58.6)	19 (27.1)	65 (29.0)	< 0.0001‡
Dysphagia, yes	54 (63.5)	32 (47.8)	116 (54.2)	0.1374
Double vision, yes	44 (50.0)	18 (25.7)	71 (31.4)	0.0018‡

SD, standard deviation; SCA, spinocerebellar ataxia; BMI, body mass index; SARA, Scale for the Assessment and Rating Ataxia; INAS, Inventory of Non-ataxia Signs.

*Means with the same letter did not differ significantly from each other according to a Tukey's Honestly Significant Difference multiple comparison test ($P > 0.05$).

†The chi-square test was used to analyze qualitative covariates, and an analysis of variance was used for quantitative covariates.

‡ $P < 0.05$ (statistically significant).

§Values indicate the length of the expanded allele (repeat units).

greater frequency of number of non-ataxia signs (INAS score, 5 ± 2 signs), especially pyramidal and peripheral motor symptoms: spasticity (44%), hyperreflexia (47%), extensor plantar (41%), muscle atrophy (39%), fasciculation (43%), and brainstem oculomotor signs (59%). Paresis, rigidity, resting tremor, sensory symptoms, urinary dysfunction, cognitive impairment, dysphagia, anxiety or depression, the use of any psychiatric medication, cancer, and gastrointestinal and endocrine diseases like diabetes all were similar between groups. Moreover, we failed to identify any factors that were associated with a BMI increase. Although we did not find a correlation with dysphagia, the decreasing BMI group included 64% of patients

who had dysphagia, whereas the increasing and stable BMI groups included 54% and 48% of patients, respectively.

Trajectories of BMI and Disease Progression

The association between the BMI groups and disease progression, as determined with the SARA score, was analyzed (Fig. 2). Disease progression was significantly faster for the decreasing BMI group (SARA score, 5.7 ± 0.73 points every decade), intermediate for the stable BMI group (SARA score, 4.2 ± 0.47 points every decade), and slower for the group

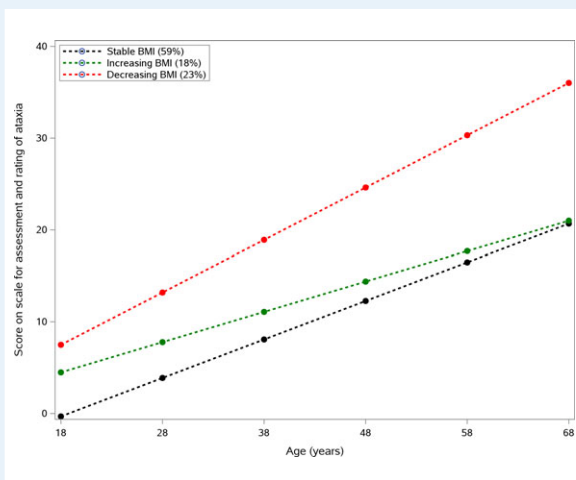


Figure 2 Progression of scores on the Scale for the Assessment and Rating of Ataxia according to body mass index (BMI) evolution. Stable BMI (in kg/m²; black curve) = $-0.32 + 4.2 \times \text{age}$; increasing BMI (green curve) = $4.5 + 3.3 \times \text{age}$; and decreasing BMI (red curve) = $7.5 + 5.7 \times \text{age}$.

increasing BMI group (SARA score 3.3 ± 0.78 points every decade).

Discussion

In this first long-term follow-up study of body weight changes in a large cohort of patients with SCA, a decline in BMI over time was observed. Furthermore, three trajectories of BMI evolution were identified: decreasing, increasing, and stable. There was a significant association between weight loss and disease progression, but no significant correlation of BMI change with the presence of dysphagia was observed.

At baseline, patients with SCA3 had a lower BMI than the other patients. These results are consistent with findings from the Brazilian cohort at the Hospital de Clinicas de Porto Alegre (HCPA), who demonstrated that that BMI was lower in patients who had SCA3 compared with controls and inversely was correlated with expanded CAG repeat lengths.⁹ In addition, these results are in line with observations in several transgenic SCA models, in which mice had progressive body weight loss over the course of the disease.^{11–14} We observed that BMI declined over time. Similar results were reported in longitudinal clinical studies of Huntington's disease over 3 years, in which BMI declined with time, and patients with higher CAG repeat numbers had faster rates of weight loss.⁸

To our knowledge, this is the first study to allow the identification of subgroups with respect to the evolution of body weight among patients with polyglutamine disorders. Unlike the LM model, which shows an average decrease BMI over time, our subgroup analysis showed that only 23% of patients lost weight during follow-up. Three groups were objectified with decreasing, increasing, and stable BMI over time. The most interesting finding was that the decreasing BMI group had faster disease progression, suggesting that weight loss may serve

as a biomarker to monitor disease progression. In addition, longer CAG expansions were associated with BMI declines in patients with SCA1, SCA2, and SCA3. This was not replicated for those with SCA6, because there was only 1 patient in the decreasing BMI group. These results reinforced the assumption of a correlation between genetic findings, disease progression, and weight loss for patients with SCA1, SCA2, and SCA3, but not for those with SCA6.

Although ataxia is the major symptom of all SCAs, various non-ataxia signs may accompany ataxia. We observed a greater frequency of non-ataxia signs (especially motor symptoms) in patients who had declining BMI over time. On average, each patient in the group with decreasing BMI had 5 non-ataxia symptoms. From our results, muscle atrophy is very likely to contribute to total weight loss, because 39% of patients who had decreasing BMI over time had muscle atrophy. This frequency increases to 75% in patients with stage 3 disease. In line with that finding, a significant correlation was observed between disease stage and BMI: 45% of patients who were permanently wheelchair dependent had a declining BMI over time. By contrast, no such association was reported in the Brazilian cohort study.⁹

The rates of progression in our study were slower than the values reported in a previous analysis of the same cohort.² This discrepancy is explained by the different time scales used in the analyses. Here, we analyzed data on an age scale, as recommended for chronic diseases, in which age is the main disease evolution factor.²⁰ In addition, unlike the previous report, all SCA groups were pooled for the analysis, because the sample size for each SCA group separately did not allow the fit of LCM models with more than 2 subgroups.

The mechanism underlying weight loss in patients with SCA is not clear. It has been suggested that the occurrence of dysphagia may be an important cause of weight loss.¹⁰ However, the available results are contradictory. The current findings agree with those from the Brazilian study, in which no relation of BMI with dysphagia was reported.⁹ In contrast, a systematic post-mortem study on thick serial tissue sections from 12 dysphasic patients with SCA2, SCA3, SCA6, and SCA7 reported the presence of widespread neurodegeneration of the brainstem nuclei involved in the ingestive process that may cause nutritional deficiencies, weight loss, and dehydration.¹⁰

In addition, only 12 patients (3%) from the EUROSCA cohort developed severe dysphagia during follow-up, requiring the use a tube feeding. In theory, this type of diet should increase the patient's weight. However, among those patients, there were 4 (42%) in the increasing BMI group, 5 (42%) in the decreasing BMI group, and 7 (58%) in the stable BMI group. These results suggest that severe dysphagia has no impact on increases in BMI, and it is also evident that weight loss is a primary consequence of disease progression.

The increased catabolic state in patients with SCA1 was reported as a possible mechanism involved in weight loss.¹⁵ In that study, the authors demonstrated that patients with SCA1 had a 22% increase in energy expenditure per kilogram of fat-free mass and a 28% increase in fat oxidation compared with controls. This increasing catabolic state might relate to the

presence of autonomic dysfunction, as has been suggested in HD.²¹ Several studies have reported the presence of autonomic dysfunction in patients with symptomatic SCA1, SCA2, and SCA3 who may be able to reduce parasympathetic activity.^{22,23} In addition, the lower motor neuron degeneration, which is described in SCA1, SCA2, and SCA3, but not in SCA6, may have an influence on BMI decline.²⁴ In amyotrophic lateral sclerosis, lower BMI might be caused by hypermetabolism.²⁵ It is likely that the same mechanisms are present in patients with SCA1, SCA2, and SCA3. The greater prevalence of hypermotoric signs like dystonia or dyskinesia in the group with decreasing BMI point in the same direction, because such signs have been associated with metabolic changes and weight loss in other disorders like Parkinson's disease.²⁶

Alterations in the insulin/insulin-like growth factor-1 system have been involved as another mechanism for weight loss in transgenic animal models and in data from patients with SCA3.^{27,28} In the Brazilian cohort study, investigators observed that patients with SCA3 had reduced insulin levels and that the insulin levels were directly associated with BMI.^{9,28} Similar results were reported in HD⁸ but were not replicated in all studies in which, despite high caloric intake, the weight loss took place. This allowed the use of a plasma biomarker to identify early stages of the disease.²⁹ It also allowed the development of a spectroscopic cerebral marker of brain energy metabolism, which could be improved by using triheptanoin to provide substrates to the Krebs cycle.³⁰

Our study has several strengths. First, it had a multicenter, prospective design with long observational periods. Second, to account of informative dropout, we applied a sensitivity analysis using a joint model for longitudinal data and the time to dropout.²⁰ The results obtained from this joint model were similar to estimations using standard methods, suggesting an absence of bias in parameter estimations (Table S3). Third, we used a compromise criterion to select the best model instead of using only the Bayesian Information Criterion. The average probability of a model with 3 classes was relatively high, ranging from 0.74 to 0.83, suggesting an unambiguous classification. However, the data have limitations: Information about caloric intake, lipid profile, daily exercise, physiotherapy, food habits, gait training, occupational therapy, and olfactory impairment, all of which may influence BMI, was not assessed in the EUROSCA study. In addition, other anthropometric measurements, such as skin-fold thickness or waist circumference, which could possibly serve as biomarkers to monitor disease progression, were missing; and the small sample size per genotype negated the possibility of performing a trajectory analysis for each SCA type.

The current data allowed us to identify 3 different groups of BMI evolution and provided evidence of an association between weight loss and disease progression. These findings suggest monitoring of body weight as a potentially useful biomarker of disease progression, because patients who had the greatest number of CAG repeats were more likely to lose weight. In addition, the results provide more information to help with patient selection and stratification for the design of interventional clinical trials.

Author Roles

1. Research Project: A. Conception, B. Organization, C. Execution; 2. Statistical Analysis: A. Design, B. Execution, C. Review and Critique; 3. Manuscript Preparation: A. Writing the First Draft, B. Review and Critique.

A.Diallo: 2A, 2B, 2C, 3A, 3B
 H.J.: 1A, 1B, 1C, 2C, 3B
 T.S.H.: 1B, 1C, 2C, 3B
 A.C.: 1B, 1C, 2C, 3B
 R.L.: 1B, 1C, 2C, 3B
 A.Durr: 1B, 1C, 2C, 3B
 A.B.: 1B, 1C, 2C, 3B
 P.C.: 1B, 1C, 2C, 3B
 C.Marelli: 1B, 1C, 2C, 3B
 C.Mariotti: 1B, 1C, 2C, 3B
 L.N.: 1B, 1C, 2C, 3B
 M.Panzeri: 1B, 1C, 2C, 3B
 M.R.: 1B, 1C, 2C, 3B
 A.Sobanska: 1B, 1C, 2C, 3B
 A.Sulek: 1B, 1C, 2C, 3B
 L.S.: 1B, 1C, 2C, 3B
 H.H.: 1B, 1C, 2C, 3B
 B.M.: 1B, 1C, 2C, 3B
 A.F.: 1B, 1C, 2C, 3B
 A.A.: 1B, 1C, 2C, 3B
 J.I.: 1B, 1C, 2C, 3B
 J.B.: 1B, 1C, 2C, 3B
 B.P.v.d.W.: 1B, 1C, 2C, 3B
 D.T.: 1B, 1C, 2C, 3B
 S.B.: 1B, 1C, 2C, 3B
 M.Pandolfo: 1B, 1C, 2C, 3B
 J.B.S.: 1B, 1C, 2C, 3B
 P.B.: 1B, 1C, 2C, 3B
 P.G.: 1B, 1C, 2C, 3B
 L.B.: 1B, 1C, 2C, 3B
 M.H.P.: 1B, 1C, 2C, 3B
 J.S.K.: 1B, 1C, 2C, 3B
 T.K.: 1A, 1B, 1C, 2A, 2C 3B
 S.T.d.M.: 1A, 1B, 1C, 2A, 2C, 3A 3B

Acknowledgments

We thank the Fondation pour la Recherche Médicale for the financial support of Alhassane Diallo. We also thank Fatoumata Diallo from East Carolina University for critical reading of the article; and Thomas Klopstock (Department of Neurology, University of Munich, Munich, Germany), Jens Petersen (Department of Neurology, University of Munich, Munich, Germany, Department of Neurology, University of Zurich, Zurich, Switzerland), Catherine Delnooz (Department of Neurology, Radboud University Nijmegen Medical Center, Nijmegen, Netherlands), and Rafal Rola (First Department of Neurology, Institute of Psychiatry and Neurology, Warsaw, Poland) for their contribution of patients and help with patient assessments.

Disclosures

Ethical Compliance Statement: We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this work is consistent with those guidelines.

Funding Sources and Conflict of Interest: The sponsors of this study had no role in design, data collection, data analysis, data interpretation, or writing the report. The corresponding author had full access to all data in the study and had final responsibility for the decision to submit for publication. The EU FP6 (EUROSCA), the German Ministry of Education and Research (GeneMove), the Polish Ministry of Science, and the EU FP7 (Neuromics) supported the study. This study was funded within the framework of EU FP6 (EUROSCA). Additional funds came from the German Ministry of Education and Research (BMBF; GeneMove), the Polish Ministry of Scientific Research and Information Technology, and the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreement 2012–305121, “Integrated European-Omics Research Project for Diagnosis and Therapy in Rare Neuromuscular and neurodegenerative diseases (NEUROMICS)” and Funding by Neuromics (EU FP7).

Financial Disclosures for previous 12 months: Thomas Klockgether reports grants from EU FP6 during the conduct of the study and personal fees from Biohaven, ICON Clinical Research, and Atheneum outside the submitted work. Anna Sobanska takes part in the clinical trial “Clinical TeleNeuroforma-clinically verified home rehabilitation system for people suffering from selected neurological conditions” sponsored by The National Center for Research and Development. Sophie Tezenas du Montcel reports grants from EU FP6 during the conduct of the study. Bart P. van de Warrenburg reports grants from ZonMW, BBMRI-NL, the Gossweiler Foundation, Hersenstichting, Radboud University Medical Center, the European Union, and Bioblast Pharma and other support from the Movement Disorder Society outside the submitted work. Jörg B. Schulz reports service on the scientific advisory boards for Lundbeck, TEVA, Novartis, and Lilly; funding for travel and personal fees from GlaxoSmithKline, Merz Pharmaceuticals, Medical Tribune, Lundbeck, Pfizer, Boehringer, and Bayer; and research support from the German Ministry of Education and Research and the European Union. Peter Bauer reports a European Commission research grant (EUROSCA), during the conduct of the study and personal fees from Actelion Pharmaceuticals and Aschwil CH and personal fees from Centogene AG and Rostock GER outside the submitted work. The remaining authors report no sources of funding and no conflicts of interest.

References

- Durr A. Autosomal dominant cerebellar ataxias: polyglutamine expansions and beyond. *Lancet Neurol* 2010;9:885–894.
- Jacobi H, du Montcel ST, Bauer P, et al. Long-term disease progression in spinocerebellar ataxia types 1, 2, 3, and 6: a longitudinal cohort study. *Lancet Neurol* 2015;14:1101–1108.
- Maas RP, van Gaalen J, Klockgether T, van de Warrenburg BP. The preclinical stage of spinocerebellar ataxias. *Neurology* 2015;85:96–103.
- Jacobi H, Reetz K, du Montcel ST, et al. Biological and clinical characteristics of individuals at risk for spinocerebellar ataxia types 1, 2, 3, and 6 in the longitudinal RISCA study: analysis of baseline data. *Lancet Neurol* 2013;12:650–658.
- Seidel K, Siswanto S, Brunt ER, den Dunnen W, Korf HW, Rub U. Brain pathology of spinocerebellar ataxias. *Acta Neuropathol (Berl)* 2012;124:1–21.
- Pedroso JL, Braga-Neto P, Escorcio-Bezerra ML, et al. Non-motor and extracerebellar features in spinocerebellar ataxia type 2. *Cerebellum* 2017;16:34–39.
- Pedroso JL, Franca MC Jr, Braga-Neto P, et al. Nonmotor and extracerebellar features in Machado-Joseph disease: a review. *Mov Disord* 2013;28:1200–1208.
- Aziz NA, van der Burg JM, Landwehrmeyer GB, Brundin P, Stijnen T, EHD Study Group; Roos RA. Weight loss in Huntington disease increases with higher CAG repeat number. *Neurology* 2008;71:1506–1513.
- Saute JA, Silva AC, Souza GN, et al. Body mass index is inversely correlated with the expanded CAG repeat length in SCA3/MJD patients. *Cerebellum* 2012;11:771–774.
- Rub U, Brunt ER, Petrasch-Parwez E, et al. Degeneration of ingestion-related brainstem nuclei in spinocerebellar ataxia type 2, 3, 6 and 7. *Neuropathol Appl Neurobiol* 2006;32:635–649.
- Iizuka A, Nakamura K, Hirai H. Long-term oral administration of the NMDA receptor antagonist memantine extends life span in spinocerebellar ataxia type 1 knock-in mice. *Neurosci Lett* 2015;592:37–41.
- Jafar-Nejad P, Ward CS, Richman R, Orr HT, Zoghbi HY. Regional rescue of spinocerebellar ataxia type 1 phenotypes by 14–3–3epsilon haploinsufficiency in mice underscores complex pathogenicity in neurodegeneration. *Proc Natl Acad Sci USA* 2011;108:2142–2147.
- Hubener J, Casadei N, Teismann P, et al. Automated behavioral phenotyping reveals presymptomatic alterations in a SCA3 genetrapp mouse model. *J Genet Genomics* 2012;39:287–299.
- Boy J, Schmidt T, Wolburg H, et al. Reversibility of symptoms in a conditional mouse model of spinocerebellar ataxia type 3. *Hum Mol Genet* 2009;18:4282–4295.
- Mahler A, Steiniger J, Endres M, Paul F, Boschmann M, Doss S. Increased catabolic state in spinocerebellar ataxia type 1 patients. *Cerebellum* 2014;13:440–446.
- Schmitz-Hubsch T, du Montcel ST, Baliko L, et al. Scale for the assessment and rating of ataxia: development of a new clinical scale. *Neurology* 2006;66:1717–1720.
- Schmitz-Hubsch T, Coudert M, Bauer P, et al. Spinocerebellar ataxia types 1, 2, 3, and 6: disease severity and nonataxia symptoms. *Neurology* 2008;71:982–989.
- Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. *JAMA* 1999;282:1737–1744.
- Verbeke G, Molenberghs G. *Linear Mixed Models for Longitudinal Data*. New York: Springer Science + Business Media; 2009.
- Commenges D, Jacqmin-Gadda H. *Dynamical Biostatistical Models*. Boca Raton, FL: CRC Press; 2015.
- Aziz NA, Pijl H, Frolich M, Snel M, Streefland TC, Roelfsema F, Roos RA. Systemic energy homeostasis in Huntington's disease patients. *J Neurol Neurosurg Psychiatry* 2010;81:1233–1237.
- Netravathi M, Sathyaprabha TN, Jayalaxmi K, Datta P, Nirmala M, Pal PK. A comparative study of cardiac dysautonomia in autosomal dominant spinocerebellar ataxias and idiopathic sporadic ataxias. *Acta Neurol Scand* 2009;120:204–209.
- Montes-Brown J, Sanchez-Cruz G, Garcia AM, Baez ME, Velazquez-Perez L. Heart rate variability in type 2 spinocerebellar ataxia. *Acta Neurol Scand* 2010;122:329–335.
- Liang L, Chen T, Wu Y. The electrophysiology of spinocerebellar ataxias. *Neurophysiol Clin* 2016;46:27–34.
- Paganoni S, Deng J, Jaffa M, Cudkowicz ME, Wills AM. Body mass index, not dyslipidemia, is an independent predictor of survival in amyotrophic lateral sclerosis. *Muscle Nerve* 2011;44:20–24.
- Kashihara K. Weight loss in Parkinson's disease. *J Neurol* 2006;253(suppl 7):VII38–VII41.

27. Duarte-Silva S, Neves-Carvalho A, Soares-Cunha C, Teixeira-Castro A, Oliveira P, Silva-Fernandes A, Maciel P. Lithium chloride therapy fails to improve motor function in a transgenic mouse model of Machado-Joseph disease. *Cerebellum* 2014;13:713–727.
28. Saute JA, da Silva AC, Muller AP, et al. Serum insulin-like system alterations in patients with spinocerebellar ataxia type 3. *Mov Disord* 2011;26:731–735.
29. Mochel F, Charles P, Seguin F, et al. Early energy deficit in Huntington disease: identification of a plasma biomarker traceable during disease progression. *PLoS ONE* 2007;2:e647.
30. Adanyeguh IM, Rinaldi D, Henry PG, et al. Triheptanoin improves brain energy metabolism in patients with Huntington disease. *Neurology* 2015;84:490–495.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Table S1. Summary of the estimated latent class mixed model on the EUROSCA sample (n=384).

Table S2. Posterior classification.

Table S3. Sensitivity analysis.

Figure S1. Flow-chart.

Table S1: Summary of the estimated latent class mixed model on the EUROSCA sample (n=384):

G	L	P	BIC	%class1	%class2	%class3	%class4	%class5
1	-2925.09	13	5927.54	100				
2	-2918.64	18	5944.39	16.40	83.60			
*3	-2908.75	23	5954.37	58.85	18.23	22.92		
4	-2898.40	28	5963.42	1.30	57.55	17.97	23.18	
5	-2893.80	33	5983.98	32.81	35.16	0.78	13.02	18.23

Number of latent classes (G), log-likelihood (L), number of parameters (P), Bayesian Information Criterion (BIC), proportion in each latent class (in%).

Table S2: Posterior classification:

*Mean posterior probability to				
Final	Subjects	belong to class (%)		
Classification	(%)	Stable	Increasing	Decreasing
1	226 (58.85)	0.8324	0.1748	0.2271
2	70 (18.23)	0.0684	0.8024	0.0318
3	88 (22.92)	0.0992	0.0228	0.7411

Number (%) of subjects and mean posterior probabilities (given in %) of the latent class membership according to the final posterior classification.

*The values in bold indicate the best model with the average mean posterior probability (Table 1 and 2).

Table S3: Sensitivity analysis:

	LMM	JM	
Parameters	(MAR)	(MNAR)	Bias
Fixed-effects			
Intercept	25.938	25.889	0.048
Linear age	-3.071	-3.126	0.055
Quadratic age	1.138	1.182	-0.043
Cubic age	-0.106	-0.112	0.005
Variance-covariances B			
SD (b0)	7.573	6.879	0.693
SD (b1)	3.419	3.140	0.278
SD (b2)	0.080	0.068	0.012
Cor (b0, b1)	-0.947	-0.919	-0.027
Cor (b0, b2)	-0.328	-0.256	-0.071
Cor (b2, b3)	-0.427	-0.388	-0.039
Autoregressive correlation			
AR	0.325	0.315	0.009
σ (AR)	2.610	2.742	-0.132
Measurements error			
σ	0.527	0.527	0.000

Comparison parameters estimates (fixed-effects, variance-covariance matrix, first order auto-regressive correlation and measurements error) from the linear mixed model (LMM) under MAR assumption and those from the joint model (JM) under MNAR assumption.

Figure S1: Flow-chart

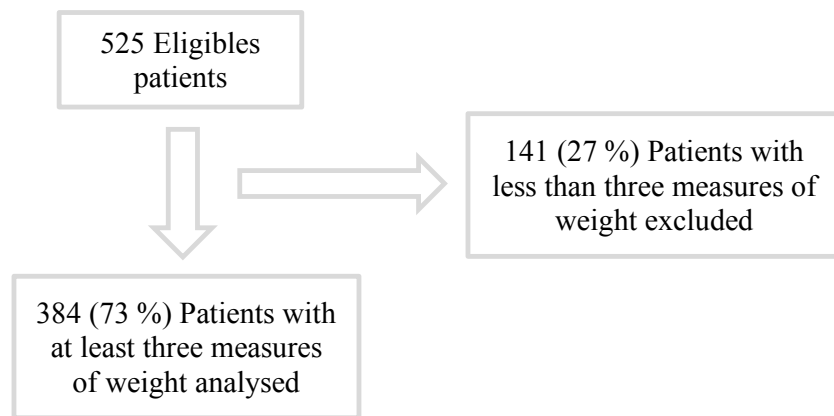
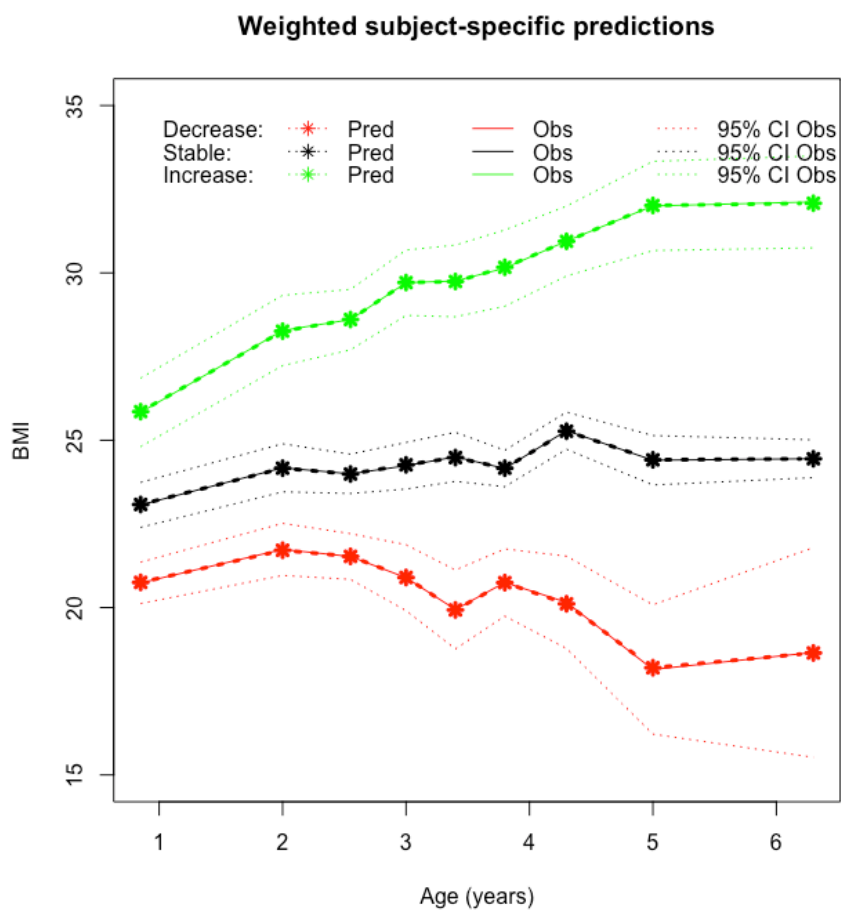


Figure S2: Weighted subject-specific predictions



5 Analyse de survie : Développement d'un Nomogramme pronostique de décès pour les patients SCA

5.1 Résumé

5.1.1 Rationnel

Bien que l'espérance de vie des patients atteints d'ataxie spinocérébelleuse soit considérablement réduite, seules quelques études [Kieling et al., 2007, Almaguer-Mederos et al., 2013, Monin et al., 2015, Klockgether et al., 1998] souffrant de quelques limites (petite taille, suivi court) ont estimé la survie ou identifié des facteurs de risque influençant la survie. Dans ce travail, notre objectif était premièrement d'étudier la survie à 10 ans des patients souffrant d'ataxie de type 1 (SCA1), SCA2, SCA3 ou SCA6 à partir de la plus grande cohorte au monde avec le plus long suivi (EUROSCA) ; secondairement d'identifier les facteurs pronostiques qui influencent cette survie ; et enfin, de développer un modèle pronostique permettant de prédire la survie individuelle en fonction des caractéristiques génétiques et cliniques.

5.1.2 Méthodes

Nous avons analysé l'ensemble des 525 patients inclus et suivis de la cohorte EUROSCA. Le critère de jugement principal était la survie globale. Pour éviter les biais de survie, nous avons mis à jour le statut vital à partir de la base de données électroniques, ou de l'interrogatoire des membres de la famille ou des mairies de naissance lorsque cela était possible. Les potentiels prédicteurs de décès ont été sélectionnés dans un premier temps parmi ceux déjà publiés dans la littérature (sexe, année de naissance, âge de début de la maladie, et CAG long) puis dans un second temps, parmi quelques variables déjà disponibles dans la cohorte EUROSCA telles que les scores cliniques (SARA et INAS), les items du score INAS et leur extension à la dysphagie, la double vision, la déficience cognitive et les troubles urinaires, le stade de la maladie, la dépression (score PHQ-9), l'utilisation de la physiothérapie, et les évaluations sur l'état de santé (poids) et de la qualité de vie (score EQ-5D VAS).

Facteurs pronostiques de décès

Dans ce chapitre, toutes les analyses ont été effectuées séparément pour chaque génotype. Nous avons défini la survie comme le délai entre l'inclusion le décès toutes causes et censuré les patients vivants ou perdus de vue à leur date de dernières nouvelles. Nous avons estimé la probabilité de

survivre pour un patient au moins jusqu'au temps t par l'estimateur du maximum de vraisemblance non-paramétrique ou méthode de Kaplan-Meier. Pour identifier les facteurs pronostiques indépendants de décès, nous avons utilisé le modèle de Cox à risques proportionnels ajustés sur l'âge à l'inclusion.

$$h_i(t, X, \beta) = h_0(t)e^{(\beta^T X)} \quad (5.1)$$

où $X = (age, sexe, \dots, scoreEQ-5DVAS)^T$ est le vecteur de m prédicteurs de décès décrit plus haut, $\beta = (\beta_{age}, \dots, \beta_{scoreEQ-5DVAS})$ le vecteur des coefficients de régression qui leur sont associés et $h_0(t)$ la fonction du risque de base. Les paramètres β ont été estimés par la vraisemblance partielle de Cox :

$$\mathcal{L}(\beta, X) = \prod_{i=1}^m P_i = \prod_{i=1}^m \frac{e^{(\beta^T X_i)}}{\sum_{m: T_k \geq t_i} e^{(\beta^T X_k)}} \quad (5.2)$$

où P_i est la probabilité conditionnelle que le patient i décède en t_i sachant qu'il est à risque de décès au temps t_i et qu'il n'y a qu'un seul décès en t_i ; m étant le nombre de facteurs.

Développement du nomogramme

A partir de chaque modèle final, nous avons construit un nomogramme incluant tous les facteurs de risque indépendants de décès pour prédire la probabilité de survie après 5 et 10 ans. Cette probabilité de survie ($t = [5 \text{ et } 10 \text{ ans}]$) a été définie par :

$$S(t, X_i) = S_0(t)e^{(LP_i)} \quad (5.3)$$

où LP_i est la fonction prédictive linéaire correspondant à : $LP_i = \sum_{i=1}^m \beta X_{im}$.

Le nomogramme est un système de points qui attribue à chaque prédicteur de décès des points compris entre 0 et 100 sur une interface graphique. La construction du nomogramme se fait en 2 étapes : l'identification du facteur le plus influent sur le risque d'évènement en l'occurrence ici le décès, et l'attribution du nombre de points à chaque facteur en fonction de son influence sur le risque de décès.

Identification du facteur le plus influent sur le risque d'évènement : pour chaque facteur selon qu'il soit binaire ou continu, un β maximum en valeur absolue est calculé selon [Iasonos et al., 2008] :

- facteur m binaire, le β maximum correspond à la valeur absolue du coefficient $|\hat{\beta}|$ directement issu du modèle de Cox : $\beta_{max_{binaire}} = |\hat{\beta}_{binaire}|$
- facteur m quantitatif, le β maximum est estimé par : $\beta_{max_{quantif}} = |\hat{\beta}_{quantif}| \times (max_{quantif} - min_{quantif})$.

Attribution des points : une fois le facteur le plus influent identifié ($|\beta_{max}|$), un nombre de points lui est ainsi attribué.

- Si le facteur avec le $|\beta_{max}|$ est quantitatif, on lui affecte 0 point pour la valeur minimale et 100 points pour la valeur maximale.

- Si le facteur avec le $|\beta_{max}|$ est binaire, on lui affecte 0 point pour la modalité codé 0 et 100 points pour la modalité codé 1.

Pour les autres facteurs moins influents m , le nombre de points est calculé par :

- $point_m = 100 \times \left(\frac{|\beta_m|}{|\beta_{max}|} \right)$

d'où le nombre total de point est : $TotalPoints = \sum_m point_m$

Ensuite, nous avons calculé un score pronostique en additionnant la contribution de chaque facteur individuel selon que le facteur est binaire ou continu, en fonction des points qui lui sont attribués par le nomogramme $|\beta_{max}|$:

$$Score_{nomogramme} = (Quantim - min_{Quantim}) \times \left(\frac{|\beta_{max}|}{|max_{Quantim} - min_{Quantim}|} \right) + Binaire_m \binom{0}{1} \times (|\beta_{max}|) + \dots \quad (5.4)$$

Puis, nous avons regroupé le score dérivé du nomogramme en 3 catégories selon la distribution suivante : 55% pour le groupe de bon pronostique, 25% pour le groupe de pronostique intermédiaire et 20% pour le groupe de mauvais pronostique. Enfin, nous avons évalué la capacité discriminative du nomogramme par le $c - index$ et l'adéquation aux données par la calibration.

5.1.3 Résultats

Les résultats de ce travail ont permis de confirmer que la survie était différente selon le génotype. La survie était plus courte pour les SCA1, intermédiaire pour les SCA2 et SCA3, et plus longue pour les SCA6. Nous avons identifié pour chaque génotype des facteurs de risques indépendants de décès. Ce sont la présence d'une dysphagie et un score SARA élevé pour les SCA1, l'âge avancé et du nombre de répétitions du CAG long, et un score SARA élevé pour les SCA2, la présence d'une dystonie, un score SARA élevé et l'interaction négative entre l'âge et le CAG pour les SCA3, et un score SARA élevé pour les SCA6. Les nomogrammes que nous avons développés à partir des facteurs de risque indépendant de décès avaient de très bonnes performances en termes de discrimination et de calibration, mais nécessitent une validation externe avant d'être utilisés en pratique clinique courante. De plus, ces nomogrammes pourraient aider à faciliter la sélection et la stratification des patients pour les futurs essais cliniques.

L'ensemble de ce travail a fait l'objet d'une publication dans le Lancet Neurology.

5.2 Article 2 : The Lancet Neurology



Survival in patients with spinocerebellar ataxia types 1, 2, 3, and 6 (EUROSCA): a longitudinal cohort study

Alhassane Diallo, Heike Jacobi, Arron Cook, Robyn Labrum, Alexandra Durr, Alexis Brice, Perrine Charles, Cecilia Marelli, Caterina Mariotti, Lorenzo Nanetti, Marta Panzeri, Maria Rakowicz, Anna Sobanska, Anna Sulek, Tanja Schmitz-Hübsch, Ludger Schöls, Holger Hengel, Bela Melegh, Alessandro Filla, Antonella Antenora, Jon Infante, José Berciano, Bart P van de Warrenburg, Dagmar Timmann, Sylvia Boesch, Massimo Pandolfo, Jörg B Schulz, Peter Bauer, Paola Giunti, Jun-Suk Kang, Thomas Klockgether, Sophie Tezenas du Montcel

Summary

Background Spinocerebellar ataxias are dominantly inherited progressive ataxia disorders that can lead to premature death. We aimed to study the overall survival of patients with the most common spinocerebellar ataxias (SCA1, SCA2, SCA3, and SCA6) and to identify the strongest contributing predictors that affect survival.

Methods In this longitudinal cohort study (EUROSCA), we enrolled men and women, aged 18 years or older, from 17 ataxia referral centres in ten European countries; participants had positive genetic test results for SCA1, SCA2, SCA3, or SCA6 and progressive, otherwise unexplained, ataxias. Survival was defined as the time from enrolment to death for any reason. We used the Cox regression model adjusted for age at baseline to analyse survival. We used prognostic factors with a p value less than 0.05 from a multivariate model to build nomograms and assessed their performance based on discrimination and calibration. The EUROSCA study is registered with ClinicalTrials.gov, number NCT02440763.

Findings Between July 1, 2005, and Aug 31, 2006, 525 patients with SCA1 (n=117), SCA2 (n=162), SCA3 (n=139), or SCA6 (n=107) were enrolled and followed up. The 10-year survival rate was 57% (95% CI 47–69) for SCA1, 74% (67–81) for SCA2, 73% (65–82) for SCA3, and 87% (80–94) for SCA6. Factors associated with shorter survival were: dysphagia (hazard ratio 4.52, 95% CI 1.83–11.15) and a higher value for the Scale for the Assessment and Rating of Ataxia (SARA) score (1.26, 1.19–1.33) for patients with SCA1; older age at inclusion (1.04, 1.01–1.08), longer CAG repeat length (1.16, 1.03–1.31), and higher SARA score (1.15, 1.10–1.20) for patients with SCA2; older age at inclusion (1.44, 1.20–1.74), dystonia (2.65, 1.21–5.53), higher SARA score (1.26, 1.17–1.35), and negative interaction between CAG and age at inclusion (0.994, 0.991–0.997) for patients with SCA3; and higher SARA score (1.17, 1.08–1.27) for patients with SCA6. The nomogram-predicted probability of 10-year survival showed good discrimination (c index 0.905 [SD 0.027] for SCA1, 0.822 [0.032] for SCA2, 0.891 [0.021] for SCA3, and 0.825 [0.054] for SCA6).

Interpretation Our study provides quantitative data on the survival of patients with the most common spinocerebellar ataxias, based on a long follow-up period. These results have implications for the design of future interventional studies of spinocerebellar ataxias; for example, the prognostic survival nomogram could be useful for selection and stratification of patients. Our findings need validation in an external population before they can be used to counsel patients and their families.

Funding European Union 6th Framework programme, German Ministry of Education and Research, Polish Ministry of Scientific Research and Information Technology, European Union 7th Framework programme, and Fondation pour la Recherche Médicale.

Introduction

Spinocerebellar ataxias are a clinically and genetically heterogeneous group of dominantly inherited autosomal progressive ataxia disorders. Several genetically distinct spinocerebellar ataxias have been defined, the most common being SCA1, SCA2, SCA3, and SCA6. These subtypes are caused by translated CAG repeat expansions that encode elongated polyglutamine tracts within the various proteins associated with each type. In addition, there are other spinocerebellar ataxias caused by non-translated repeat expansions or conventional mutations. Clinically, these spinocerebellar ataxias are characterised by progressive balance problems and incoordination

(with onset most common during adulthood), which lead to severe disability and premature death.¹

Although life expectancy of patients with SCA1, SCA2, SCA3, and SCA6 is substantially reduced, only a few studies have investigated survival or identified factors that affect survival. Significantly decreased survival associated with CAG repeat number, age at onset, and year of birth was recorded in a study of patients with SCA3 in Brazil.² Similar findings were obtained for Cuban patients with SCA2, except for the effect of year of birth on survival.³ A study of 446 index cases with known SCA mutations in spinocerebellar ataxia genes and 509 affected relatives reported a lower age of death in

Lancet Neurol 2018; 17: 327–34

See [Comment](#) page 292

Institut Pierre Louis d'Epidémiologie et de Santé Publique IPLESP (A Diallo MD, S Tezenas du Montcel PhD), Institut du cerveau et la moelle épinière (ICM) (Prof A Durr MD, Prof A Brice MD), Sorbonne Université, INSERM, Paris, France; Genetics Department (P Charles MD), Assistance Publique-Hôpitaux de Paris AP-HP, Pitié-Salpêtrière University Hospital Paris, Paris, France (S Tezenas du Montcel, Prof A Durr, Prof A Brice); Department of Neurology, University Hospital of Heidelberg, Heidelberg, Germany (H Jacobi MD, T Schmitz-Hübsch MD, Prof T Klockgether); German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany (H Jacobi, T Schmitz-Hübsch, Prof T Klockgether, L Schöls, H Hengel MD); Department of Molecular Neuroscience, Institute of Neurology (A Cook MD, P Giunti PhD), Neurogenetics Laboratory, National Hospital of Neurology and Neurosurgery (R Labrum PhD), University College London, London, UK; Service de Neurologie—CMRR, CHRU Gui de Chauiac, Montpellier, France (C Marelli MD); SOSD Genetics of Neurodegenerative and Metabolic Diseases, Fondazione-IRCCS Istituto Neurologico Carlo Besta, Milan, Italy (C Mariotti MD, L Nanetti MD, M Panzeri MD); Department of Clinical Neurophysiology (M Rakowicz PhD, A Sobanska MD), Department of Genetics (A Sulek PhD), Institute of Psychiatry and Neurology, Warsaw, Poland; Charité-Universitätsmedizin Berlin, NeuroCure Clinical Research Center, Clinical Neuroimmunology Group,

Berlin, Germany (T Schmitz-Hübsch MD); Department of Neurodegeneration and Hertie-Institute for Clinical Brain Research (L Schöls MD, H Hengel), Institute of Medical Genetics and Applied Genomics (P Bauer MD), University of Tübingen, Tübingen, Germany; German Center for Neurodegenerative Diseases (DZNE), Tübingen, Germany (L Schöls, H Hengel); Department of Medical Genetics, and Szentagothai Research Center, University of Pécs, Pécs, Hungary (Prof B Melegh PhD); Department of Neuroscience, and Reproductive and Odontostomatological Sciences, Federico II University Naples, Italy (Prof A Filla MD, A Antenora MD); Service of Neurology, University Hospital Marqués de Valdecilla (IDIVAL), University of Cantabria (UC), Santander, Spain (J Infante MD, Prof J Berciano MD); Centro de Investigación Biomédica en Red de Enfermedades Neurodegenerativas (CIBERNED), Santander, Spain (J Infante, Prof J Berciano); Department of Neurology, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, Nijmegen, Netherlands (B P van de Warrenburg PhD); Department of Neurology, Essen University Hospital, University of Duisburg-Essen, Essen, Germany (D Timmann MD); Department of Neurology, Medical University Innsbruck, Innsbruck, Austria (S Boesch MD); Université Libre de Bruxelles (ULB), Neurology Service, ULB Hôpital Erasme, ULB Laboratory of Experimental Neurology, Brussels, Belgium (Prof M Pandolfo MD); Department of Neurology, RWTH Aachen University, Aachen, Germany (Prof J B Schulz MD); JARA-Translational Brain Medicine, Aachen-Jülich, Aachen, Germany (Prof J B Schulz); Department of Neurology, Zala County Hospital, Zalaegerszeg, Hungary (Prof B Melegh PhD); Department of Neurology, University of Frankfurt, Frankfurt, Germany (J-S Kang MD); Department of

Research in context

Evidence before this study

We searched PubMed using the search terms ["spinocerebellar ataxia" OR "dominant ataxia" OR "Machado-Joseph disease" AND "survival"] for reports published before Oct 31, 2017. Only peer-reviewed, English-language reports of studies done in patients were considered. We identified four studies: two (Almaguer-Mederos LE, 2013, and Kielsing C, 2007) were restricted to patients with mutations in one SCA gene (SCA2 or SCA3); the third (Monin M-L, 2015) focused on index cases with known mutations in SCA genes and their parents (if affected), and analysed spinocerebellar ataxias due to polyglutamine expansions as a group. The fourth (Klockgether T, 1998) was an international retrospective study with short follow-up. These studies were heterogeneous in terms of design, follow-up time, and methods. Few predictors of death were identified (sex, year of birth, age at onset, and repeat lengths of the expanded allele).

Added value of this study

In this European, multicentre, longitudinal study (EUROSCA), we prospectively investigated a large cohort of patients with

patients with polyglutamine expansions than in those with other types of mutation.⁴ Additionally, among the patients with polyglutamine expansions, survival was significantly shorter for patients with SCA1 than for those with SCA2, SCA3, SCA6, or SCA7.⁴

EUROSCA is a European multicentre longitudinal cohort study of patients with SCA1, SCA2, SCA3, and SCA6. It was initiated in 2005 with the goal of characterising the natural history of the disease and identifying prognostic factors. We recorded phenotypic differences between patients with the different disease subtypes at baseline and identified factors that determined disease severity. Analyses of longitudinal data after 2 years⁵ and 8 years⁶ allowed us to establish disease subtype-specific progression rates and identify factors that predict the course of the disease. In this study, we report survival data of the EUROSCA participants based on an observational period of 10 years. This study had three aims: to quantify overall survival of patients with SCA1, SCA2, SCA3, or SCA6; to identify prognostic factors that influence survival; and, finally, to develop a prognostic model that allows prediction of individual survival based on genetic and clinical characteristics.

Methods

Study design and population

The study population consisted of patients with spinocerebellar ataxia recruited from 17 ataxia referral centres in ten European countries in the longitudinal EUROSCA prospective cohort study,⁷ between July 1, 2005, and Aug 31, 2006. Patients were identified with the help of an electronic patient registry that contains data for all patients with spinocerebellar ataxias who had been in

SCA1, SCA2, SCA3, or SCA6 for 10 years. Overall survival rates differed with respect to disease subtype. A higher Scale for the Assessment and Rating of Ataxia score at baseline was associated with shorter survival for all disease subtypes. This study highlighted for the first time the effect of additional signs (dysphagia or dystonia) on survival. It is the first to address a development of prognostic survival model in ataxia patients allowing a precise prediction of survival on an individual basis.

Implications of all available evidence

The available data provide quantitative information about survival of patients with SCA1, SCA2, SCA3, and SCA6, and allowed the identification of predictors of survival. The prognostic nomograms could help researchers to optimise the design of future clinical trials, but need validation in an external populations before they can be used to assist clinicians in counselling patients and their families.

contact with one of the study centres. These patients had progressive, otherwise unexplained, ataxia and had a positive molecular genetic test for SCA1, SCA2, SCA3, or SCA6. Assessments were done according to a written study protocol. Patients were seen at baseline and followed up by annual visits for 3 years. Afterwards, participants entered an extension phase in which assessments were done during routine visits, resulting in irregular intervals between the visits. The database was locked on Nov 3, 2016, after a maximum observation period of 11 years. The ethics committees of the participating centres approved the study. Written informed consent was obtained from all study participants at enrolment.

Outcome and predictor variables

The clinical outcome was overall survival. To avoid survival bias, we retrieved the updated vital status from the electronic database when available, through interviews of family members, and by interrogation of records from civil registry offices when feasible. As candidate predictors, we selected sex, age at onset, and repeat lengths of the expanded alleles, which have been reported as predictors of death in previously published studies.²⁻⁴ As additional candidates, we selected disease duration, age, and factors that characterise the neurological phenotype (as measured by the Scale for the Assessment and Rating of Ataxia [SARA], Inventory of Non-Ataxia Signs [INAS], and individual non-ataxia signs), mood (Patient Health Questionnaire-9 [PHQ-9]),⁸ and physical state (BMI and disease stage) of the study participants. Any use of physiotherapy was included because it is the only known therapeutic intervention in ataxia. The appendix provides the complete list of candidate

predictors (appendix p 5). BMI at baseline was calculated using the formula (weight/height²). Scores on SARA (range 0–40),⁷ total scores on the INAS (0–16), and individual non-ataxia signs as given in the INAS, including reported abnormalities such as dysphagia and double vision,⁹ were recorded at baseline. The PHQ-9 is a nine item self-rating questionnaire that simply scores each of the nine DSM-IV criteria for depressive disorders. The severity of depression is calculated by assigning scores of 0, 1, 2, and 3, to the response categories “not at all”, “several days”, “more than half the days” and “nearly every day”, respectively. The sum score ranges from 0 (absence of depression) to 27 (severe depression). Repeat lengths of the expanded and normal alleles were determined at the Institute of Medical Genetics and Applied Genomics of the University of Tübingen (Tübingen, Germany).

Statistical analysis

We used mean (standard deviation) or frequencies (percentages) to describe the continuous and categorical variables at baseline. Survival was calculated from the date of enrolment to death for any reason. Data for patients who were alive or lost to follow-up were censored. We used time from enrolment as the timescale. Survival was estimated using the Kaplan-Meier method and compared with a log-rank test. Cox-proportional-hazard models adjusted for age at baseline were used to study prognostic factors, and then candidates with a *p* value less than 0.10 were entered into the multivariate Cox regression. The strongest contributing predictors for death from multivariate regression were selected through a backward procedure based on the lowest Akaike information criterion. Because the backward selection model relies on statistical significance, which in turn depends on statistical power, we computed a posteriori statistical power to highlight an increase in the risk of death of 2 (hazard ratio [HR] 2) for any binary variable with a proportion of patients of 50% in each group. We did subsequent analyses separately for each disease subtype as survival differed between the subtypes. Assumptions of hazard proportionality and log-linearity were verified.

We constructed a nomogram that included the selected prognostic factors from each final Cox model to estimate the probability of survival after 5 and 10 years. A raw prognostic score was computed by summing the contribution of each individual factor, based on the points given for each factor in the nomogram. We divided the patients of each disease subtype into three prognostic risk groups to provide a reasonable spread of risk. Thus, various cutoffs for the risk score were explored based on the optimal cutoff using the three-risk group approach.¹⁰ The distribution of 55% for prognostic risk group 1 (good prognosis), 25% for prognostic group 2 (intermediate prognosis), and 20% for prognostic group 3 (poor prognosis), was used, as these percentages were the most

discriminative relative to the *c* index (appendix pp 4, 6). We did calibration plots and computed the discrimination *c*-index to assess the performance of the nomograms. 1000 random samples of the population were used to derive the 95% CI bootstrap percentile for the *c*-index.^{11,12} The multivariate model was internally validated using the 1000 samples bootstrap procedure.

All data analyses were done with SAS (version 9.4) and R (version 3.3.3). We deemed *p* values lower than 0.05 to be statistically significant; all tests were two-sided.

Role of the funding source

The funders of the study had no role in the design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Results

Between July 1, 2005, and Aug 31, 2006, we enrolled 525 patients with SCA1 (*n*=117), SCA2 (*n*=162), SCA3 (*n*=139), or SCA6 (*n*=107). Table 1 shows the clinical and demographic data. During the follow-up, 66 (13%) patients were lost to follow-up and 121 (23%) died: 36 (31%) with SCA1, 38 (24%) with SCA2, 34 (25%) with SCA3, and 13 (12%) with SCA6. Causes of death were reported for 57 (47%) of 121 patients. Among them, pulmonary diseases (pneumonia or pulmonary insufficiency) were the most common (six in SCA1, four in SCA2, seven in SCA3, and one in SCA6), followed by unknown causes (three in SCA1, three in SCA2, six in SCA3, and two in SCA6), cardiac causes (five in SCA3 and three in SCA6), cancer (one in SCA1, two in SCA2, and three in SCA6), sepsis or cachexia (one in SCA1, one in SCA2, and two in SCA6), gastric causes (two in SCA2), suicide (one in SCA1 and one in SCA2), stroke (one in SCA2 and one in SCA3), and renal disease (one in SCA2).

The 5-year survival rate was 80% (95% CI 72–88) for SCA1, 87% (82–93) for SCA2, 87% (82–93) for SCA3, and 98% (95–100) for SCA6. The corresponding 10-year survival rates were 57% (95% CI 47–69), 74% (67–81), 73% (65–82), and 87% (80–94). Overall, survival was significantly different between the subtypes of spinocerebellar ataxia (figure 1). The risk of death was higher for patients with SCA1, SCA2, and SCA3 than those with SCA6 (figure 1). Moreover, the risk of death in patients with SCA1 was higher than in those with SCA2 (HR 1.80, 95% CI 1.14–2.8) or SCA3 (1.63, 1.02–2.61), whereas it was similar between patients with SCA2 or SCA3 (0.91, 0.57–1.44).

We applied univariate Cox regression modelling adjusted for age at baseline to identify predictors for death for each disease subtype. For SCA1, the 19 (68%) of 28 evaluated predictors associated with survival (appendix p 5) were CAG repeat number, age at onset, disease duration, BMI, disease stage, PHQ-9 sum score, SARA score, number of non-ataxia signs, and several individual

Neurology, University Hospital of Bonn, Bonn, Germany (Prof T Klockgether, L Schöls MD, H Hengel MD)

Correspondence to: Dr Sophie Tezenas du Montcel, Sorbonne Université, Institut Pierre Louis d'Epidémiologie et de Santé Publique IPLESP, F75013 Paris, France
sophie.tezenas@aphp.fr

See Online for appendix

	Patients with SCA1 (n=117)	Patients with SCA2 (n=162)	Patients with SCA3 (n=139)	Patients with SCA6 (n=107)
Sex				
Men	71 (61%)	74 (46%)	73 (53%)	58 (54%)
Women	46 (39%)	88 (54%)	66 (47%)	49 (46%)
Age (years)	46.3 (12.2)	46.4 (13.3)	48.8 (11.8)	65.0 (10.9)
Age at onset (years)	36.7 (10.4)	35.2 (12.5)	37.0 (11.3)	54.3 (10.6)
Disease duration (years)	9.5 (5.7)	11.1 (6.0)	11.7 (6.0)	10.7 (6.9)
Number of CAG repeats	47.6 (5.6)	39.4 (3.5)	69.1 (4.6)	22.5 (2.2)
BMI (kg/m ²)	24.4 (4.1)	25.1 (4.3)	23.1 (4.0)	25.7 (4.1)
SARA score	15.6 (9.1)	15.7 (8.0)	15.1 (8.5)	15.2 (6.8)
PHQ-9 sum score	6.7 (6.4)	5.5 (4.9)	6.9 (6.2)	5.3 (5.4)
Any physiotherapy use (yes)	43 (37%)	75 (46%)	75 (54%)	51 (48%)
Number of non-ataxia signs	4.8 (2.2)	4.2 (2.2)	4.9 (2.6)	1.9 (1.6)
Death (yes)	36 (31%)	38 (24%)	34 (25%)	13 (12%)
Median follow-up (years)	9.9 (6.9–10.1)	10.2 (10.1–10.3)	10.2 (10.1–10.4)	10.2 (10.1–10.3)

Data are n (%) mean (SD), or median (95% CI). SARA=Scale for the Assessment and Rating of Ataxia. SCA=spinocerebellar ataxia. PHQ-9=Patient Health Questionnaire-9.

Table 1: Population characteristics at enrolment

impairment (appendix p 5). The strongest contributing risk factors for death (table 2) were older age at inclusion, longer CAG repeat number, and higher SARA score.

For SCA3, the significant predictors in the univariate analysis were CAG repeat number, age at onset, disease duration, disease stage, PHQ9 sum score, SARA score, number of non-ataxia signs, and the following individual non-ataxia signs: extensor plantar signs, spasticity, paresis, muscle atrophy, fasciculation, rigidity, chorea or dyskinesia, dystonia, and brainstem oculomotor (appendix p 5). The strongest contributing risk factors (table 2) were older age at inclusion, dystonia, higher SARA score, and a negative interaction between CAG repeat number and age at inclusion. Patients with CAG-expanded alleles with less than 62 repeats had a higher risk of death for older age at inclusion, whereas those with more than 62 repeats had a higher risk of death for younger age at inclusion, independent of the severity of the disease. In addition, there was an age-dependent effect on the risk of death associated with CAG-expanded alleles: the older the age at baseline, the weaker the effect of CAG-expanded alleles on the risk of death, and conversely, the younger the age at baseline, the stronger the effect.

In SCA6, disease stage, SARA score, and rigidity were significantly associated with survival in Cox univariate analyses (appendix p 5). The only contributing risk factor for death was a higher SARA score (table 2).

The nomograms to predict the probability of 5-year and 10-year survival for each disease subtype (appendix) showed that the SARA score contributed the most strongly to the prognosis for SCA1, SCA2, and SCA6 (appendix p 1), whereas age at baseline and its interaction with CAG were the strongest factors for SCA3 (appendix p 2). The number of CAG repeats and age at baseline had a moderate effect on the survival of patients with SCA2, whereas dysphagia and dystonia had little effect on the survival of patients with SCA1 and SCA3, respectively. A raw score was computed from the nomograms, and the patients were classified into three risk categories: good, intermediate, and poor prognosis based on the thresholds (defined in the appendix p 6). As an example of use of the nomograms and classification of patients as having good, intermediate, or poor prognosis, a 50-year-old patient with SCA1 (3 points), dysphagia (17 points), and a SARA score of 20 (50 points) has 70 points, placing him in the intermediate group with a probability of 5-year survival of 80% (95% CI 73–87) and 10-year survival of 5% (1–9; appendix p 1).

The prognostic nomograms had good discriminatory capacity for all SCAs (appendix p 7). The adjusted *c* indices for predicting death at 10 years were 0.905 (SD 0.027) for SCA1, 0.822 (0.032) for SCA2, 0.891 (0.021) for SCA3, and 0.825 (0.054) for SCA6. Moreover, the Kaplan-Meier curves of the three stratification risk groups were clearly separated (figure 2). The calibration plots showed excellent agreement between the nomogram prediction and the actual predicted probability of 5-year and 10-year survival

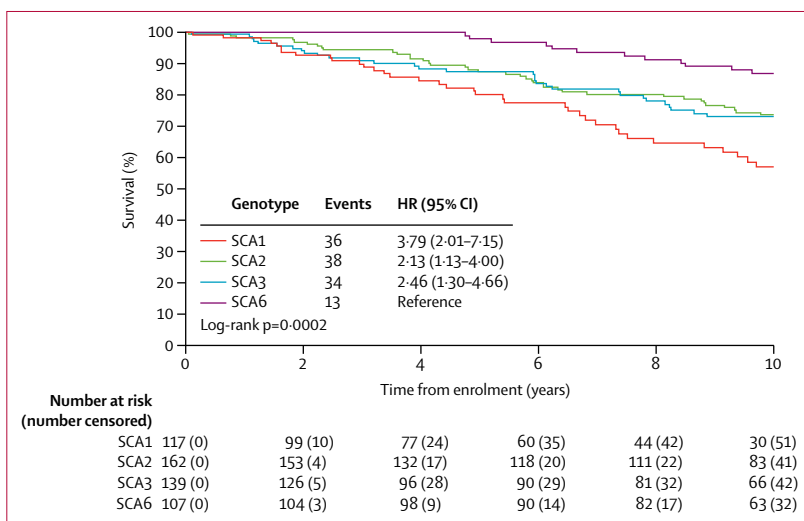


Figure 1: Survival from enrolment according to disease subtype
HR=hazard ratio. SCA=spinocerebellar ataxia.

non-ataxia signs: areflexia, paresis, muscle atrophy, fasciculation, myoclonus, chorea or dyskinesia, dystonia, resting tremor, urinary dysfunction, cognitive impairment, and dysphagia (appendix p 5). A predictive model obtained from the multivariate Cox analysis, adjusted on age at baseline identified two strongest contributing risk factors for death (table 2): dysphagia and higher SARA score.

For SCA2, the significant predictors in the univariate analysis were CAG repeat number, age at onset, disease duration, disease stage, SARA score, number of non-ataxia signs, and the following individual non-ataxia signs: paresis, muscle atrophy, fasciculation, myoclonus, rigidity, chorea, or dyskinesia, dystonia and cognitive

	HR, 95% CI	p value	Median (IQR)	Maximum score in nomogram*	Internal validation (BHR, 95% BCI)
SCA1					
Age at baseline (years)	1.008, 0.982–1.035	0.5302	46.0 (37.0–55.0)	6	1.003, 0.975–1.033
Dysphagia (yes)	4.518, 1.831–11.152	0.0011	0 (1.0–1.0)	17	7.029, 1.939–41.81
SARA score	1.256, 1.186–1.33	<0.0001	13.5 (8.5–20.5)	100	1.288, 1.201–1.382
SCA2					
Age at baseline (years)	1.042, 1.009–1.076	0.0130	47.5 (37.0–54.0)	51	1.045, 1.009–1.082
CAG length (number repeats)	1.162, 1.029–1.313	0.0158	39.0 (37.0–41.0)	53	1.172, 1.028–1.337
SARA score	1.151, 1.102–1.203	<0.0001	14.0 (10.0–19.5)	100	1.155, 1.103–1.210
SCA3					
Age at baseline (years)	1.441, 1.196–1.737	0.0001	48.0 (40.0–56.0)	100	1.692, 1.179–2.486
Dystonia (yes)	2.652, 1.208–5.534	0.0151	0	4	2.898, 1.132–7.341
SARA score	1.256, 1.173–1.346	<0.0001	14.0 (10.0–20.5)	33	1.295, 1.189–1.406
CAG (number repeats)	1.036, 0.888–1.209	0.6501	69.0 (66.0–72.0)	5	1.151, 0.867–1.533
Interaction age with CAG	0.994, 0.991–0.997	<0.0001	3328.0 (2850.0–3776.0)	86	0.993, 0.988–0.998
SCA6					
Age at baseline (years)	1.016, 0.952–1.084	0.6426	67.0 (58.0–73.0)	14	1.018, 0.947–1.096
SARA score	1.173, 1.081–1.274	0.0001	14.0 (10.5–19.0)	100	1.187, 1.080–1.306

The table shows results that were significant predictors of survival at $p < 0.05$ in the multivariate Cox model, plus age at baseline for each spinocerebellar ataxia subtype whether it was significant or not, for reference; the exception is inclusion of CAG repeat length for SCA3, as this relates to the significant predictor of the interaction of CAG repeat length with age. HR=hazard ratio. BHR=bootstrap hazard ratio. BCI=bootstrap confidence interval. SARA=Scale for the Assessment and Rating of Ataxia. SCA=spinocerebellar ataxia.* Maximum number for the highest recorded value attributed by a nomogram for each predictor.

Table 2: Multivariate Cox model

(appendix p 4). The uncertainties measured by the bootstrapping procedure in the internal validation were close to the estimated HR, except for a slight deviation for the dysphagia parameter in patients with SCA1 (table 2), suggesting robustness of the final model.

In an a posteriori calculation, the statistical power to highlight an increased risk of death (HR=2) for any binary variable with a proportion of patients of 50% in each group was 55% for SCA1, 58% for SCA2, 53% for SCA3, and 24% for SCA6.

Discussion

In this study, we have identified predictors and provided estimates of survival of patients with SCA1, SCA2, SCA3, or SCA6, based on 10-year longitudinal findings of the EUROSCA cohort.

Strengths of our study include the large number of patients and the prospective design. The 10-year observation period might seem short in view of an estimated survival of patients with spinocerebellar ataxia of 20–30 years after onset;⁴¹³ however, the average disease duration of participants at inclusion was about 10 years. Thus, we followed up the patients to an average of 20 years after ataxia onset. The information about vital status in the study database was incomplete because most of the deceased patients had not attended the study centres in the last years before death. We therefore used alternative approaches: interviews of family members and records from civil registry offices, to update the information on vital status. This allowed

us to recover 40% (49/121) of the deaths and reduced the bias due to censorship, which is one of the major sources of bias in survival analyses. Other possible sources of bias were the omission of a relevant covariate and the missing covariate being a confounder.¹⁴ We were unable to consider all potential covariates in our model. For example, clinical signs other than those assessed by the scales used in the study, imaging data, and biomarker data were not available.

In survival analysis, there is an ongoing debate about the choice of the optimal timescale. In this study, we used time-on-study rather than disease duration or chronological age as the timescale. Various simulation studies^{15–17} have shown the time-on-study scale, adjusted for age at inclusion, to be the best because it is the most robust against mis-specification (small bias) and is more suitable than other timescales to measure predictive discrimination, such as the time-dependent area under the curve. We compared the different time scales to choose the most appropriate for our data. Our empirical results agreed with the simulation studies in that time-on-study was the most accurate timescale (unpublished data).

Survival was shortest for SCA1, intermediate for SCA2 and SCA3, and longest for SCA6. These results corroborate the finding reported by Monin and colleagues,⁶ that the age at death was lower in SCA1 than in other spinocerebellar ataxias caused by polyglutamine expansions. Together, these and our findings characterise SCA1 as having the least favourable prognosis among this group of spinocerebellar ataxias. Correspondingly, two longitudinal

studies found that the progression of ataxia severity in SCA1 was faster than in SCA2, SCA3, or SCA6.^{5,6,18}

The 10-year death rate of the patients with SCA2 in our cohort (24%) was lower than that of a Cuban SCA2 cohort (29%).³ Similarly, the death rate for patients with SCA3 (25%) was lower than that of a Brazilian cohort (35%).² Possible reasons for these discrepancies are selection bias, genetic background, and differences between health-care systems.

We used Cox regression modelling to identify predictors for death in each disease subtype. For patients with SCA1, ataxia severity measured with SARA score and the presence of dysphagia increased the risk of death. These findings are contrary to the results of an

international retrospective study.¹³ However, that study used different timescales and did not include clinical findings in the statistical models. For SCA2, the risk of death increased with older age at inclusion, longer CAG repeat number, and severity of ataxia. Longer CAG repeats were similarly found to be a risk factor for death in a Cuban SCA2 cohort.³ We were unable to confirm the effect of being female on death that was reported in the international retrospective study.¹³ For SCA3, predictors of shorter survival identified by univariate Cox analysis, such as early age of onset and long CAG repeat length, overlapped with factors reported in the Brazilian SCA3 study.² None of the previous studies reported predictors for the risk of death in patients with SCA6. We found

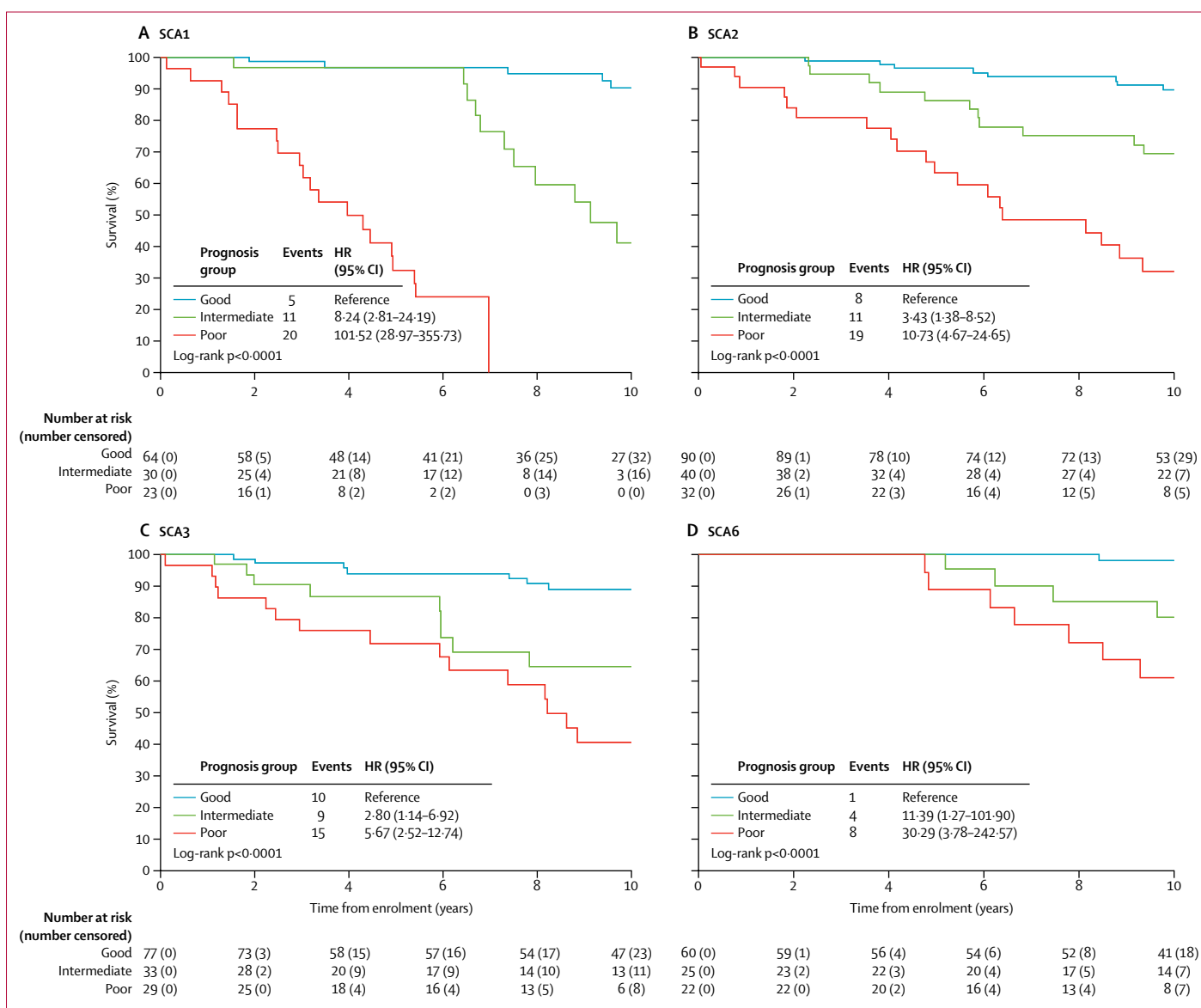


Figure 2: Survival from enrolment according to the characterisation score from nomograms by disease subtype
HR=hazard ration. SCA=spinocerebellar ataxia.

that only the severity of ataxia at baseline measured with SARA affected survival. The finding was a predictor of survival in all disease subtypes underlines the clinical relevance and predictive power of SARA. In our analysis, we failed to find an effect of physiotherapy on survival, although previous studies had shown a temporary symptomatic effect.^{19,20} We cannot rule out that we might have missed some risk factors owing to low study power.

Nomograms are widely used for prognosis in various fields of medicine. For example, nomograms allow prediction of lymph node metastasis in patients with cancer. These nomograms can assist physicians in decisions on surgical management.^{21–23} A prognosis nomogram was developed to predict individual outcomes after antiepileptic drug withdrawal in people with epilepsy.²⁴ One main limitation of the nomogram is that it assumes that outcomes remain constant over time. Consequently, its accuracy decreases over time, probably because of changes in natural history of the disease, early diagnosis detection, and improvements in treatment.²⁵ We constructed nomograms for each disease subtype that allowed prediction of individual survival with high precision based on a number of easily accessible factors identified in the Cox model. Nomograms were constructed in a rigorous methodological framework, including the choice of the candidate predictors and timescale.¹⁶ The nomograms had good discriminatory capacity, and there was excellent agreement between the nomogram prediction and actual survival. Cox proposed use of the following distribution when categorising the prognostic score in three groups: 27%, 49·5%, and 27%.¹⁰ This distribution was not appropriate for our data in view of the large number of patients with good prognoses. We therefore chose the distribution that optimised the separation of the Kaplan-Meier curves and thus retained the distribution with the largest *c* index (appendix p 6). However, the nomograms need to be externally validated on independent samples including non-European patients to establish the generalisability of the model.^{11,12} They might be further improved by the incorporation of imaging data and biomarkers.

To conclude, our data extend the knowledge of the biological characteristics of patients with SCA1, SCA2, SCA3, and SCA6. Furthermore, our nomograms could facilitate selection and stratification of patients for future clinical trials.

Contributors

ADi designed and executed the statistical analysis, wrote the first draft of the report, and reviewed the report. HJ and TS-H contributed to the conception, organisation, and execution of the research project and reviewed and commented on the statistical analysis and the report. AC, RL, ADu, AB, PC, CeM, CaM, LN, MarP, MR, ASu, LS, HH, BM, AF, AA, JI, JB, BPvdW, DT, SB, MasP, JBS, PB, PG, and J-SK organised and did the research project and reviewed and commented on the statistical analysis and the report. TK conceived, organised, and did the research project; designed, reviewed, and commented on the statistical analysis; wrote the first draft of the report, and reviewed the report. STdM contributed to the conception of the research project, wrote the first draft of the report, and reviewed the report.

Declaration of interests

STdM reports grants for the European integrated project on spinocerebellar ataxias (EUROSCA) from European Union 6th Framework programme. ADi reports grants from Fondation pour la Recherche Médicale (grant number: PLP20151034334). TK reports grants for the European integrated project on spinocerebellar ataxias (EUROSCA) from European Union 6th Framework programme and grants from the European Union Joint Programme—Neurodegenerative Disease Research (JPND) for the European Spinocerebellar Ataxia Type 3/Machado-Joseph Disease Initiative (ESMI; grant number 01ED1602A). MasP reports grants from Fonds de la recherche scientifique (FRS-FNRS) (Belgium) and Euroataxia, grants and personal fees from Biomarin and Voyager Therapeutics, and personal fees from Apopharma, Vertex, and Pfizer. MR and ASu report grants from Clinical TeleNeuroforma (grant number IS-2/230/NCBR/2015). BPvdW reports grants from Radboud University Medical Center, Hersenstichting, The Netherlands Organisation for Health Research and Development (ZonMW), Biobanking and BioMolecular resources Research Infrastructure The Netherlands (BBMRI-NL), Gossweiler Foundation, and Bioblast Pharma. PB reports grants for EUROSCA from European Union 6th Framework programme, and personal fees from Centogene AG and Actelion Pharmaceuticals. TS-H reports grants for EUROSCA from European Union 6th Framework programme and grants from Bundesministerium für Wirtschaft und Energie and Ipsen Pharma. JBS reports grants for the German Network of Hereditary Movement Disorders (GeNeMove) from the German Federal Ministry for Education and Research (BMBF), grants for the European integrated project on spinocerebellar ataxias (EUROSCA) from European Union 6th Framework programme, and grants from the European Union (grant number HEALTH-F2-2010-242193). He serves on scientific advisory boards for Lundbeck Inc, TEVA, Novartis, ForwardPharma, and Lilly, has received funding for travel and speaker honoraria from GlaxoSmithKline, Merz Pharmaceuticals, Medical Tribune, Lundbeck Inc, Pfizer Inc, Boehringer, and Bayer, and serves as Editor-in-Chief of the *Journal of Neurochemistry* and Associate Editor for *eNeuro*. JS-K reports personal fees from Merz Pharma, Medtronic, and Ipsen Pharma. JB, AA, AF, DT, HH, LS, RL, AC, ASu, HJ, SB, JI, CaM, LN, MarP, BM, ADu, AB, CeM, PC, JB, and PG declare no competing interests.

Acknowledgments

This study was funded within the European Union 6th Framework programme (EUROSCA). Additional funds came from the German Federal Ministry of Education and Research (BMBF; GeneMove), Polish Ministry of Scientific Research and Information Technology, and the European Union's 7th Framework Programme (FP7/2007-2013) under grant agreement number 2012-305121 "Integrated European-omics research project for diagnosis and therapy in rare neuromuscular and neurodegenerative diseases (NEUROMICS)". We thank Thomas Klopstock (Department of Neurology, University of Munich, Munich, Germany), Jens Petersen (Department of Neurology, University of Munich, Munich, Germany, Department of Neurology, University of Zurich, Zurich, Switzerland), Cathérine Delnooz (Department of Neurology, Maxima Medical Centre, Veldhoven, Netherlands), and Rafał Rola (First Department of Neurology, Institute of Psychiatry and Neurology, Warsaw, Poland) for contribution of patients and help in patient assessment; and the Fondation pour la Recherche Médicale for the financial support of Alhassane Diallo, grant number PLP20151034334.

References

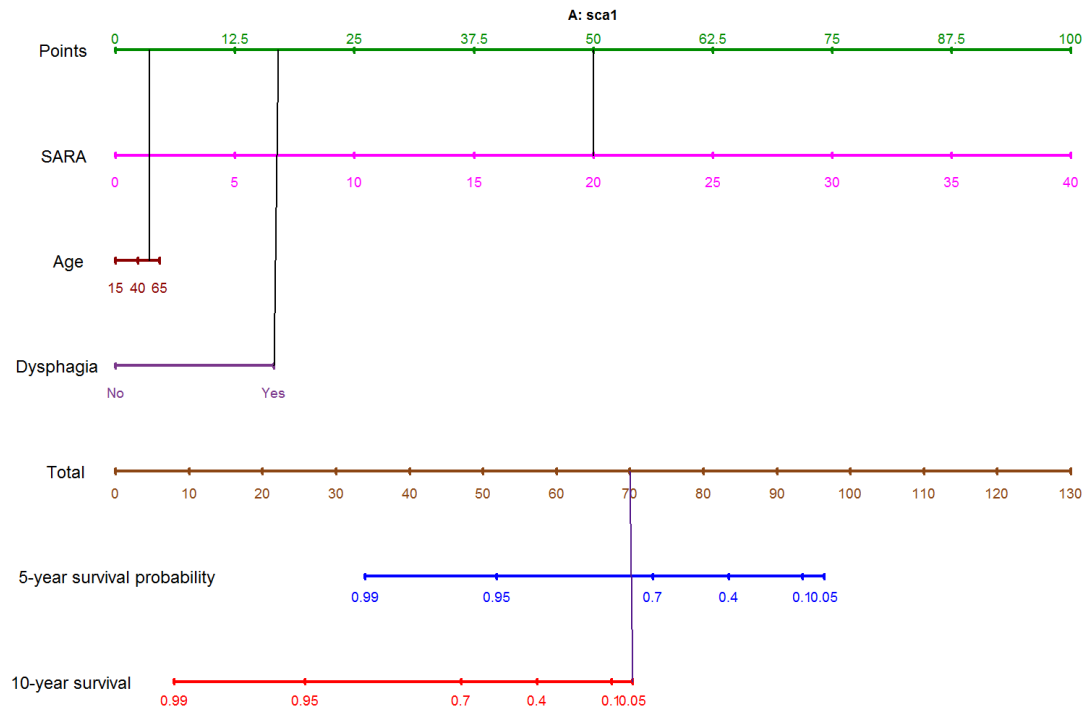
- 1 Durr A. Autosomal dominant cerebellar ataxias: polyglutamine expansions and beyond. *Lancet Neurol* 2010; **9**: 885–94.
- 2 Kielling C, Prestes PR, Saraiva-Pereira ML, Jardim LB. Survival estimates for patients with Machado-Joseph disease (SCA3). *Clin Genet* 2007; **72**: 543–45.
- 3 Almaguer-Mederos LE, Aguilera Rodríguez R, González Zaldivar Y, et al. Estimation of survival in spinocerebellar ataxia type 2 Cuban patients. *Clin Genet* 2013; **83**: 293–94.
- 4 Monin M-L, Tezenas du Montcel S, Marelli C, et al. Survival and severity in dominant cerebellar ataxias. *Ann Clin Transl Neurol* 2015; **2**: 202–07.
- 5 Jacobi H, Bauer P, Giunti P, et al. The natural history of spinocerebellar ataxia type 1, 2, 3, and 6: a 2-year follow-up study. *Neurology* 2011; **77**: 1035–41.

- 6 Jacobi H, du Montcel ST, Bauer P, et al. Long-term disease progression in spinocerebellar ataxia types 1, 2, 3, and 6: a longitudinal cohort study. *Lancet Neurol* 2015; **14**: 1101–08.
- 7 Schmitz-Hübsch T, du Montcel ST, Baliko L, et al. Scale for the assessment and rating of ataxia: development of a new clinical scale. *Neurology* 2006; **66**: 1717–20.
- 8 Spitzer RL, Kroenke K, Williams JB. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. *JAMA* 1999; **282**: 1737–44.
- 9 Schmitz-Hübsch T, Coudert M, Bauer P, et al. Spinocerebellar ataxia types 1, 2, 3, and 6: disease severity and nonataxia symptoms. *Neurology* 2008; **71**: 982–89.
- 10 Cox DR. Note on grouping. *J Am Stat Assoc* 1957; **52**: 543–47.
- 11 Royston P, Altman DG. External validation of a Cox prognostic model: principles and methods. *BMC Med Res Methodol* 2013; **13**: 33.
- 12 van Houwelingen HC. Validation, calibration, revision and combination of prognostic survival models. *Stat Med* 2000; **19**: 3401–15.
- 13 Klockgether T, Lüdtke R, Kramer B, et al. The natural history of degenerative ataxia: a retrospective study in 466 patients. *Brain J Neurol* 1998; **121**: 589–600.
- 14 Lin NX, Logan S, Henley WE. Bias and sensitivity analysis when estimating treatment effects from the cox model with omitted covariates. *Biometrics* 2013; **69**: 850–60.
- 15 Chalise P, Chicken E, McGee D. Baseline age effect on parameter estimates in Cox models. *J Stat Comput Simul* 2012; **82**: 1767–74.
- 16 Chalise P, Chicken E, McGee D. Performance and prediction for varying survival time scales. *Commun Stat - Simul Comput* 2013; **42**: 636–49.
- 17 Chalise P, Chicken E, McGee D. Time scales in epidemiological analysis: an empirical comparison. *Int J Stat Probab* 2016; **5**: 91.
- 18 Ashizawa T, Figueroa KP, Perlman SL, et al. Clinical characteristics of patients with spinocerebellar ataxias 1, 2, 3 and 6 in the US; a prospective observational study. *Orphanet J Rare Dis* 2013; **8**: 177.
- 19 Keller JL, Bastian AJ. A home balance exercise program improves walking in people with cerebellar ataxia. *Neurorehabil Neural Repair* 2014; **28**: 770–78.
- 20 Synofzik M, Ilg W. Motor Training in degenerative spinocerebellar disease: ataxia-specific improvements by intensive physiotherapy and exergames. *BioMed Res Int* 2014; **2014**: 583507.
- 21 Kim D-Y, Shim S-H, Kim S-O, et al. Preoperative nomogram for the identification of lymph node metastasis in early cervical cancer. *Br J Cancer* 2014; **110**: 34–41.
- 22 Hayashi Y, Xiao L, Suzuki A, et al. A nomogram associated with high probability of malignant nodes in the surgical specimen after trimodality therapy of patients with oesophageal cancer. *Eur J Cancer* 2012; **48**: 3396–404.
- 23 Thompson AM, Turner RM, Hayen A, et al. A preoperative nomogram for the prediction of ipsilateral central compartment lymph node metastases in papillary thyroid cancer. *Thyroid* 2014; **24**: 675–82.
- 24 Lamberink HJ, Otte WM, Geerts AT, et al. Individualised prediction model of seizure recurrence and long-term outcomes after withdrawal of antiepileptic drugs in seizure-free patients: a systematic review and individual participant data meta-analysis. *Lancet Neurol* 2017; **16**: 523–31.
- 25 Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol* 2015; **16**: e173 –80.

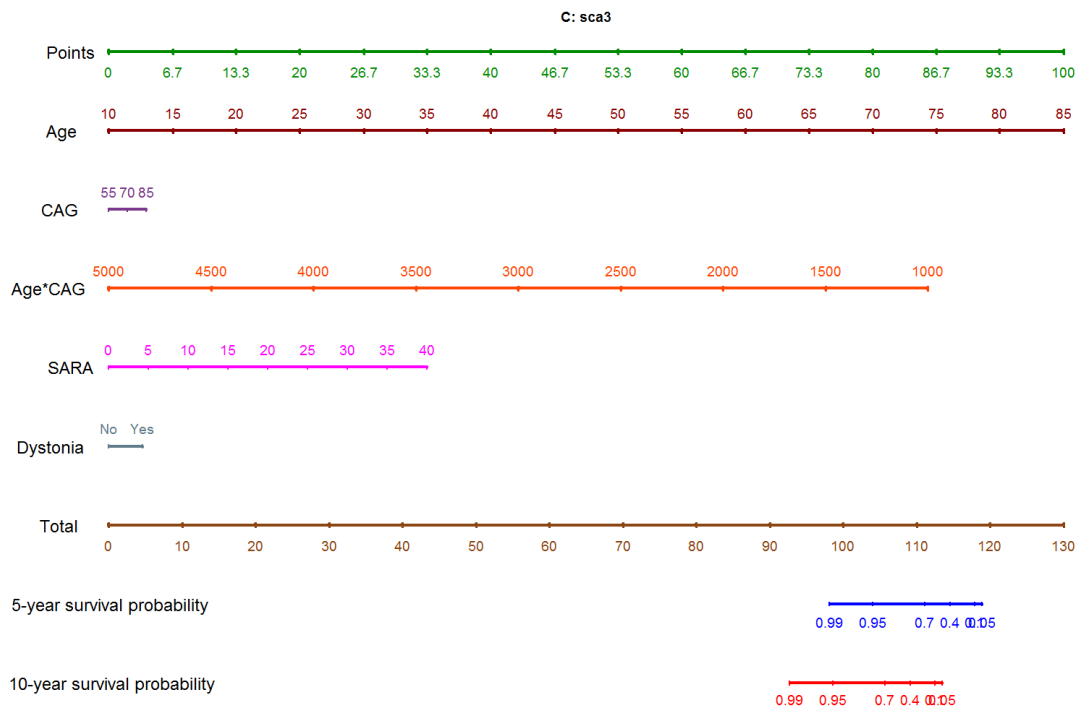
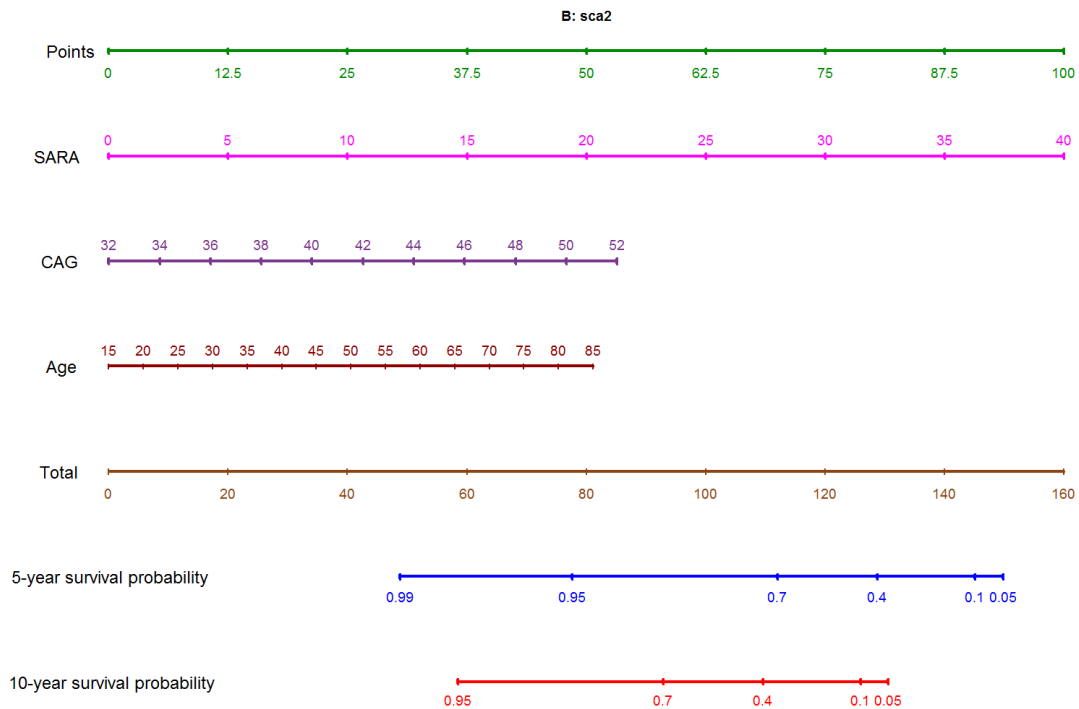
Appendix

Figure 1: Prognostic nomograms to predict the probability of individual overall survival of patients with spinocerebellar ataxia type 1, 2, 3, and 6.

Points are assigned to each risk factor by drawing a line upward from the corresponding value to the 'Points' line. The total sum point for the three factors is plotted on the 'Total points' line. A line is drawn down to read the corresponding predictions of the probability of five- and 10-year survival.



For example, a 50-year-old SCA1 patient (3 points) with dysphagia (17 points) and a SARA score of 20 (50 points) has 70 points (3 + 17 + 50), placing him in the *intermediate* group with a probability of five- (blue line) and 10-year (red line) survival of 80% and 5%, respectively.



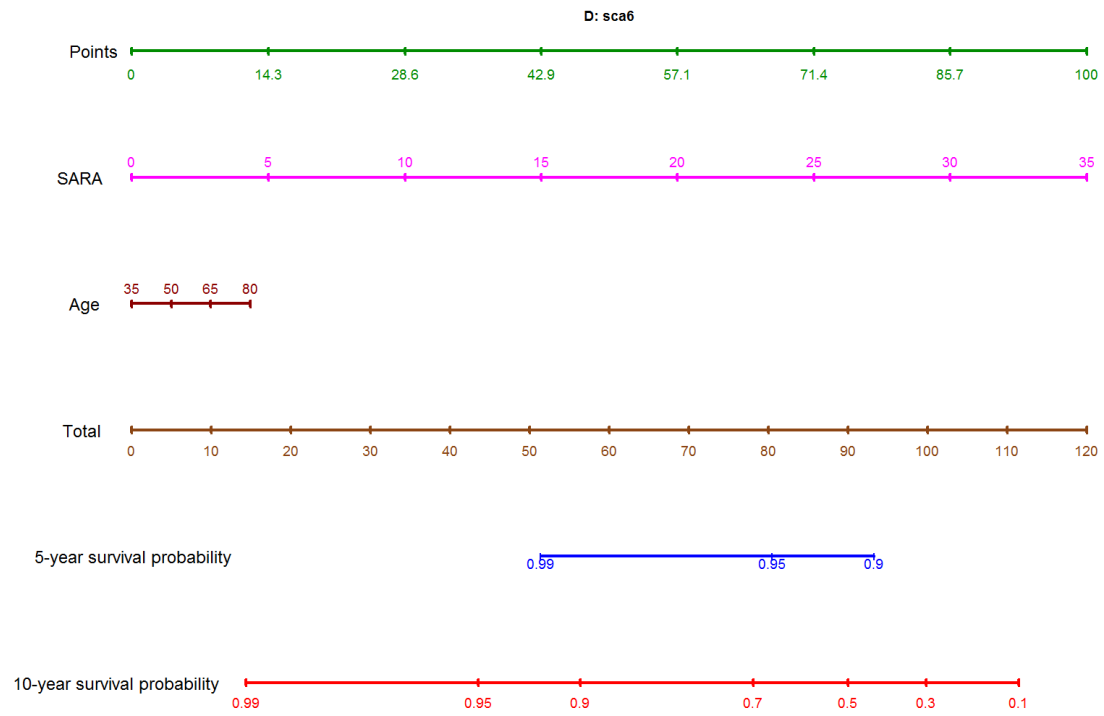
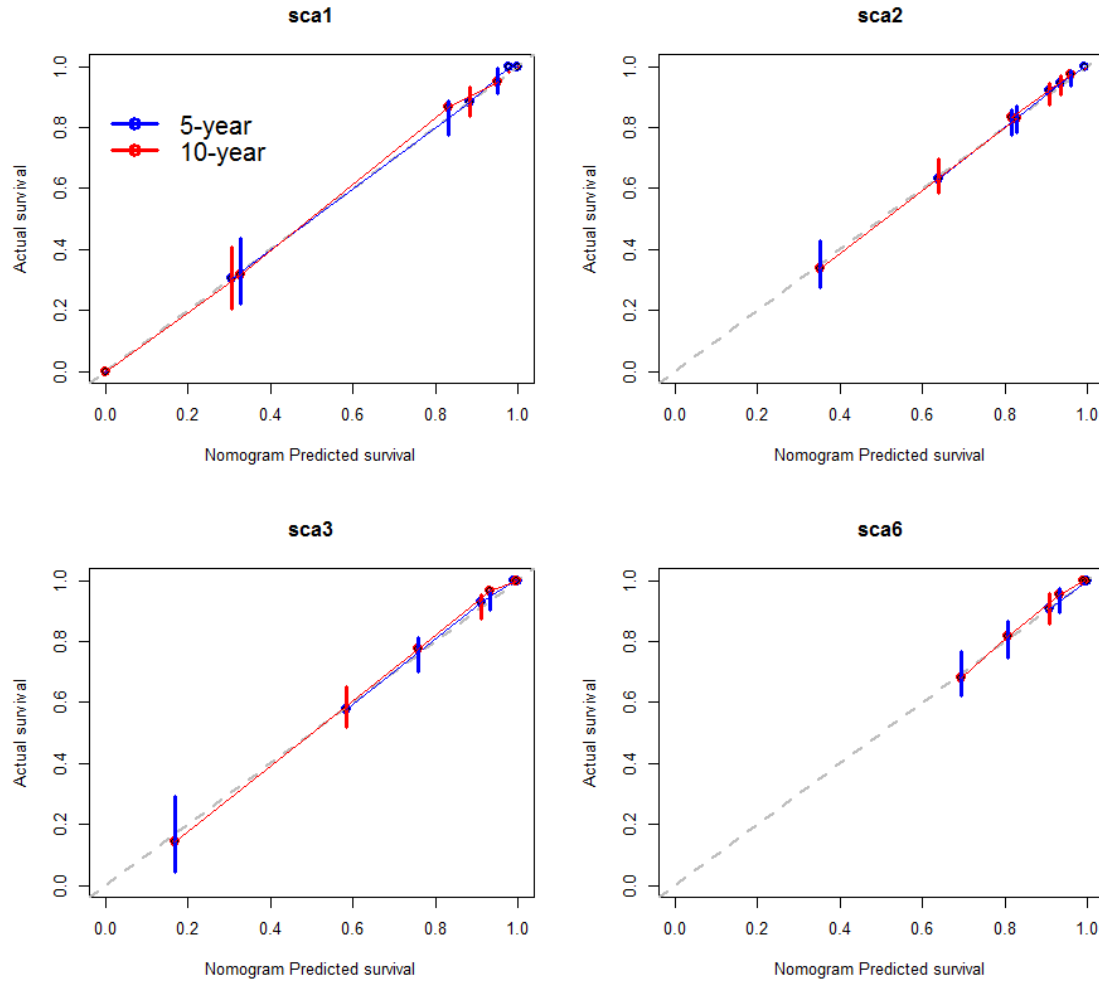


Figure 2: Calibration plots for predicting ataxia patient overall survival at each time point by genotype.



The X-axis shows the nomogram predicted probability of survival. Patients were grouped by quartiles of predicted risk. The Y-axis is the actual probability of five- (blue) and 10-year (red) survival estimated by the Kaplan-Meier method. The solid line represents the values from the nomograms with their 95% CI. A plot along a 45-degree line (dotted line) would indicate a perfect calibration model in which the predicted probabilities are identical to the actual outcomes.

Table 1: Univariate Cox model in SCAs of the 28 potential predictors of death

	SCA1	SCA2	SCA3	SCA6
	HR (95 % CI)	HR (95 % CI)	HR (95 % CI)	HR (95 % CI)
Gender (male)	0.80 (0.41 1.56)	1.34 (0.71 2.54)	1.42 (0.72 2.80)	1.61 (0.52 4.94)
Age at onset (years)	0.84 (0.78 0.90)	0.90 (0.86 0.95)	0.89 (0.85 0.94)	0.97 (0.90 1.04)
Disease duration (years)	1.20 (1.12 1.28)	1.11 (1.05 1.16)	1.12 (1.06 1.18)	1.03 (0.96 1.12)
Number of CAG repeats	1.13 (1.06 1.19)	1.27 (1.15 1.40)	1.21 (1.08 1.34)	1.11 (0.36 3.42)
BMI (kg/m ²)	0.81 (0.74 0.90)	0.94 (0.86 1.04)	0.94 (0.85 1.03)	1.02 (0.88 1.17)
SARA score	1.23 (1.17 1.29)	1.16 (1.12 1.21)	1.16 (1.12 1.20)	1.17 (1.08 1.27)
Number of non-ataxia signs	1.60 (1.32 1.95)	1.42 (1.24 1.63)	1.99 (1.62 2.46)	1.22 (0.94 1.58)
Hyperreflexia (yes)	0.66 (0.34 1.30)	0.74 (0.26 2.10)	1.73 (0.84 3.57)	0.71 (0.15 3.28)
Areflexia (yes)	2.53 (1.22 5.26)	1.82 (0.86 3.85)	1.28 (0.58 2.82)	0.75 (0.20 2.89)
Extensor plantor (yes)	2.01 (0.94 4.29)	1.21 (0.60 2.43)	2.78 (1.34 5.77)	*
Spasticity (yes)	0.67 (0.34 1.30)	1.59 (0.61 4.10)	4.14 (1.77 9.67)	2.34 (0.64 8.59)
Paresis (yes)	2.90 (1.47 5.74)	3.32 (1.60 6.91)	4.36 (2.16 8.80)	3.19 (0.71 14.42)
Muscle atrophy (yes)	3.96 (2.03 7.73)	6.14 (3.14 12.00)	2.89 (1.41 5.95)	0.45 (0.06 3.49)
Fasciculations (yes)	3.39 (1.67 6.88)	1.93 (1.00 3.69)	3.04 (1.54 6.02)	*
Myoclonus (yes)	3.22 (1.13 9.17)	2.80 (1.34 5.87)	1.84 (0.43 7.86)	*
Rigidity (yes)	2.83 (0.36 22.05)	5.44 (2.37 12.49)	3.19 (1.36 7.46)	8.83 (2.64 29.55)
Chorea/dyskenia (yes)	3.01 (1.04 8.73)	3.83 (1.46 10.02)	7.83 (3.57 17.20)	2.54 (0.31 20.80)
Dystonia (yes)	3.63 (1.73 7.61)	2.31 (1.09 4.90)	4.10 (2.08 8.07)	1.98 (0.25 15.61)
Resting tremor (yes)	3.08 (1.17 8.07)	1.27 (0.53 3.04)	2.06 (0.49 8.67)	4.38 (0.56 34.19)
Sensory symptoms (yes)	1.68 (0.74 3.82)	0.63 (0.31 1.26)	1.48 (0.63 3.49)	0.89 (0.29 2.76)
Urinary dysfunction (yes)	2.01 (1.02 3.96)	1.25 (0.65 2.38)	1.81 (0.87 3.75)	2.99 (0.96 9.29)
Cognitive impairment (yes)	3.03 (1.520 6.03)	2.63 (1.36 5.08)	1.03 (0.44 2.40)	3.13 (0.93 10.57)
Brainstem oculomotor signs (yes)	1.63 (0.85 3.14)	1.36 (0.71 2.61)	3.01 (1.40 6.46)	2.32 (0.68 7.90)
Dysphagia (yes)	4.14 (1.81 9.48)	1.59 (0.82 3.07)	1.93 (0.92 4.06)	0.98 (0.33 2.93)
Double vision (yes)	1.71 (0.77 3.84)	1.22 (0.58 2.60)	1.39 (0.69 2.81)	1.94 (0.65 5.79)
[§] Any physiotherapy use (yes)	1.85 (0.96 3.57)	1.23 (0.65 2.34)	0.79 (0.40 1.56)	0.96 (0.32 2.90)
PHQ-9 sum score	1.05 (1.00 1.10)	0.99 (0.93 1.07)	1.06 (1.02 1.11)	1.05 (0.96 1.14)
Disease stage (Independent as reference)	1	1	1	1
Dependent on walking aids	5.36 (2.07 13.83)	2.83 (1.22 6.67)	15.33 (1.99 118.21)	7.911 (0.88 57.73)
Dependent on wheelchair	48.0 (16.16 142.70)	10.25 (4.75 22.13)	91.76 (12.0 700.7)	13.64 (1.30 143.0)

*The model did not converge due to the absence of an event (death) in one of the modalities, HR: hazard ratio, CI: confidence interval, bold indicates the significance of a predictor of risk of death (HR not included 1);

[§]variable was used as time-dependent covariate in Cox model. All analyses were adjusted on age at inclusion.

Table 2: Various range and threshold prognostic score computed from nomograms

	Observed total score range	Good prognosis	Intermediate prognosis	Poor prognosis	c-index
Our proposed cut-off (55, 25 and 20%)					
SCA1	2·7 – 120·4	<48 (64) [§]	48 – 70 (29)	>70 (24)	0.905
SCA2	31·7 – 146·4	<74 (90)	74 – 102 (40)	>102 (32)	0.822
SCA3 [†]	39·9 – 131·7	<79 (77)	79 – 92 (33)	>92 (29)	0.891
SCA6	3·5 – 112·0	<53 (60)	53 – 68 (25)	>68 (22)	0.825
According to tertile (33, 33 and 33%)					
SCA1	2·7 – 120·4	<35 (39)	35 – 56 (39)	>56 (39)	0.823
SCA2	31·7 – 146·4	<62 (53)	62 – 83 (53)	>83 (55)	0.781
SCA3 [†]	39·9 – 131·7	<69 (44)	69 – 84 (44)	>84 (46)	0.715
SCA6	3·5 – 112·0	<42 (35)	42 – 58 (35)	>58 (37)	0.726
Proposed by DR Cox (27, 49.5 and 27%)					
SCA1	2·7 – 120·4	<29 (32)	29 – 45 (57)	>45 (28)	0.825
SCA2	31·7 – 146·4	<59 (44)	59 – 71 (79)	>71 (38)	0.774
SCA3 [†]	39·9 – 131·7	<65 (36)	65 – 91 (66)	>91 (32)	0.712
SCA6	3·5 – 112·0	<39 (29)	39 – 67 (53)	>67 (25)	*

[†]Interaction between age and CAG repeat length (age x CAG) was divided by 1,000 in the total score computation; *Model did not converged; [§]numbers in brackets are the number of patients in each group.

Table 3: Discrimination measures and β estimates with their SE from the Cox model

Measures	SCA1		SCA2		SCA3		SCA6	
	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
c-index	0.905	0.027	0.822	0.032	0.891	0.021	0.825	0.054
β (SE): group 2 versus 1	2.110	0.549	1.233	0.465	1.031	0.461	2.432	1.118
β (SE): group 3 versus 1	4.614	0.640	2.373	0.425	1.735	0.413	3.411	1.062

All values are based on scores from the nomograms. The risk groups were: group 1 (*Good prognosis*), group 2 (*Intermediate prognosis*), and group 3 (*Poor prognosis*). β is the estimate with their SE from the Cox model. Standard error (SE) for the c-index was estimated from 1,000 bootstrap samples.

Table 4: Individual non-ataxia signs characteristics at baseline

	SCA1 (n=117)	SCA 2 (n=162)	SCA 3 (n=139)	SCA 6 (n=107)
Hyperreflexia (n, % yes)	79 (68)	21 (13)	53 (39)	23 (22)
Areflexia (n, % yes)	21 (18)	104 (65)	78 (57)	24 (23)
Extensor plantor (n, % yes)	54 (51)	45 (31)	57 (42)	2 (2)
Spasticity (n, % yes)	67 (59)	14 (9)	60 (44)	15 (15)
Paresis (n, % yes)	27 (23)	23 (15)	34 (25)	6 (6)
Muscle atrophy (n, % yes)	34 (29)	35 (22)	52 (38)	12 (12)
Fasciculations (n, % yes)	45 (39)	61 (38)	51 (37)	3 (3)
Myoclonus (n, % yes)	5 (4)	22 (14)	6 (4)	0 (0)
Rigidity (n, % yes)	2 (2)	11 (7)	14 (10)	6 (6)
Chorea/dyskenia (n, % yes)	8 (7)	11 (7)	14 (10)	2 (2)
Dystonia (n, % yes)	15 (13)	23 (14)	33 (24)	4 (4)
Resting tremor (n, % yes)	8 (7)	24 (15)	5 (4)	2 (2)
Sensory symptoms (n, % yes)	69 (62)	106 (69)	83 (65)	49 (47)
Urinary dysfunction (n, % yes)	41 (35)	64 (40)	63 (46)	32 (30)
Cognitive impairment (n, % yes)	25 (22)	41 (26)	26 (19)	10 (9)
Brainstem oculomotor signs (n, % yes)	44 (38)	58 (36)	72 (53)	18 (17)
Dysphagia (n, % yes)	70 (60)	85 (53)	83 (60)	58 (54)
Double vision (n, % yes)	18 (15)	32 (20)	78 (56)	45 (42)

Time Scales in EUROSCA cohort Analysis: An Empirical Comparison

Table 5: Global prediction error (Brier score), C-index and Explained variation (R^2) for M1, M2 and M3. (M1: Time-on-study Model, M2: Left truncated age model, M3: Unadjusted age scale model)*

	Time-on-study Model (M1)			Left truncated age Model (M2)			Unadjusted age scale Model (M3)		
	Brier score	C-index	R^2	Brier score	C-index	R^2	Brier score	C-index	R^2
SCA1	0.330	0.896	0.181	0.710	0.683	0.040	0.708	0.541	0.039
SCA2	0.232	0.788	0.087	0.575	0.836	0.168	0.580	0.648	0.173
SCA3	0.226	0.839	0.121	0.313	0.821	0.109	0.331	0.563	0.108
SCA6	0.120	0.821	0.068	0.077	0.704	-0.007	0.087	0.502	-0.008

*All models were derived from Cox proportional hazards model with only SARA score at baseline as covariate.

Bold indicates the great value for c-index and R^2 or a small value for Brier score; For c-index and R^2 the higher the better and inversely for Brier score the lower the better.

6 Analyse conjointe de la progression de l'ataxie et de la survie

6.1 Résumé

6.1.1 Rationnel

Les analyses séparées de la progression de l'ataxie [Jacobi et al., 2015] et de la survie [Diallo et al., 2018] des patients SCA ont permis l'estimation des taux de progression et de survie spécifiques à chaque génotype et, l'identification des facteurs influençant la progression de la maladie d'une part et la survie d'autre part. Ces données ont caractérisé les SCA1 comme le génotype ayant le plus mauvais pronostic avec une progression plus rapide de la maladie et une survie plus courte, suivi des SCA2 et SCA3, et enfin des SCA6. Cependant, l'analyse conjointe de la progression de l'ataxie et de la survie présente plusieurs avantages : mieux comprendre leur interdépendance, la prise en compte de leur corrélation et l'identification de facteurs pronostiques communs. Dans ce travail, nous présentons l'étude de l'analyse conjointe de la progression de l'ataxie et de la survie. Notre objectif était de quantifier l'effet de la progression de l'ataxie sur la survie des patients SCA, d'identifier des sous-groupes de patients qui partagent la même évolution de l'ataxie et le même risque de décès, ainsi que les facteurs pronostiques influençant ces sous-groupes, et enfin prédire la survie dans une fenêtre de 5 ans basées sur la progression de l'ataxie.

6.1.2 Méthodes

Parmi les 525 patients inclus et suivis de la cohorte EUROSCA, nous avons analysé les 462 patients (107 SCA1, 146 SCA2, 122 SCA3 et 87 SCA6) ayant au moins 2 visites. Les critères de jugement clinique étaient la survie globale et la progression de la maladie. Le statut vital était le même que celui de l'étude de survie précédente. Nous avons utilisé le score SARA comme critère pour évaluer la progression de l'ataxie. Pour la sélection des facteurs prédicteurs de survie et de progression de la maladie, nous nous sommes restreints à ceux déjà publiés dans les deux études séparées à partir de la cohorte EUROSCA (sexe, âge à l'inclusion, âge de début de la maladie, CAG long, dysphagie, dystonie, SARA et INAS à l'inclusion). Nous avons défini la survie comme le délai depuis l'inclusion jusqu'au décès toutes causes et censuré les patients vivant ou perdu de vue à leur date de dernières nouvelles.

Impact de la progression de l'ataxie sur le risque de décès

Dans ce travail, nous avons utilisé la modélisation conjointe d'une part pour explorer l'association entre l'évolution du score SARA et le risque de décès, et d'autre part pour prédire le

risque de décès en fonction des mesures répétées du SARA. Pour répondre à ces questions, nous avons utilisé les deux types de modèles conjoints que nous avons décrit dans le chapitre 3 : le modèle conjoint à effets aléatoires partagés [Rizopoulos, 2012] et le modèle conjoint à classe latente [Proust-Lima et al., 2014, Commenges and Jacqmin-Gadda, 2016].

Pour quantifier et capturer la meilleure dépendance entre la survie et la progression de l'ataxie à travers le score SARA, nous avons utilisé le modèle conjoint à effets aléatoires partagés. Dans ce modèle, l'évolution du SARA en fonction du temps t a été décrit par un modèle linéaire mixte avec des effets aléatoires sur l'intercept et la pente linéaire pour tenir compte de la variabilité inter-patients. Ainsi, pour le patient i à la mesure j , le modèle s'écrit :

$$SARA_{ij} = X_{Yi}^T \beta + b_{0i} + b_{1i} \times t_{ij} + \epsilon_{ij} \quad (6.1)$$

avec

$$b_i = (b_{0i}, b_{1i})^T \sim N \left(0, B = \begin{bmatrix} \sigma_{b_0}^2 & \sigma_{b_{01}} \\ \sigma_{b_{01}} & \sigma_{b_1}^2 \end{bmatrix} \right)$$

$$\text{et } \epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$$

Le risque de décès a été décrit par un modèle à risque proportionnel :

$$h_i(t|b_i, \beta, \alpha, \alpha^{(a)}) = h_0(t) \exp(W_i^T \alpha + \kappa(b_i, t_{ij})^T \alpha^{(a)}) \quad (6.2)$$

où $h_i(t)$ est la fonction du risque de décès, $h_0(t)$ la fonction paramétrique du risque de base *Weibull* avec $h_0(t) = \gamma \lambda^\gamma t^{\gamma-1}$ avec γ la forme du risque de base, λ l'échelle. α est le vecteur de paramètres associés au vecteur de variables explicatives W_i que sont l'âge à l'inclusion, l'âge de début de la maladie, le CAG long, la dysphagie et la dystonie. La fonction $\kappa_i(b_i, t)$ définit la structure de dépendance entre le risque de décès et l'évolution du SARA et $\alpha^{(a)}$ quantifie la mesure de l'association entre les deux processus.

Afin d'explorer au mieux la dépendance entre Y_{ij} et T_i , nous avons considéré plusieurs structures de dépendances [Rizopoulos, 2012] :

- **(a) La valeur courante du score SARA** : $\kappa_i(b_i, t) = Y_i(t) = X_{Yij}^T \beta + b_{0i} + b_{1i} \times t_{ij}$; ce modèle est le plus couramment utilisé. Il suppose que le risque de décès au temps t dépend de la valeur du SARA en t débarrassée de l'erreur de mesure ;
- **(b) La pente de score SARA** : $\kappa_i(b_i, t)^T = Y_i(t)'$ où $Y_i(t)'$ est la dérivée de $Y_i(t)$ par rapport à t en t . Il suppose que le risque de décès au temps t dépend de la pente du SARA ;
- **(c) La valeur courante et la pente de SARA** : $Y_i(t), Y_i(t)'$; ce modèle est plus flexible que les deux précédents et suppose que le risque de décès au temps t dépend à la fois de la valeur courante du SARA et de sa pente.
- **(d) La valeur à l'inclusion** : $\kappa_i(b_i, t) = Y_i(0)$ où le risque de décès dépend de la valeur du SARA à l'inclusion. ;
- **(e) La variation par rapport à l'inclusion** : $\kappa_i(b_i, t) = D_i(t)$ où $D_i(t) = Y_i(t) - Y_i(0)$; le risque de décès au temps t dépend de la variation du SARA par rapport à l'inclusion ;
- **(f) La valeur à l'inclusion et la variation par rapport à l'inclusion** : $\kappa_i(b_i, t) = [D_i(t), Y_i(0)]$; ce modèle combine les modèles (d) et (e) où le risque de décès au temps t dépend à la fois de la valeur de SARA à l'inclusion et de la variation du SARA par rapport à l'inclusion ;

- **(g) Les effets aléatoires b_i :** $\kappa_i(b_i, t) = b_i$; ce modèle suppose que le risque de décès au temps t dépend des effets aléatoires à la fois sur la valeur à l'origine (b_{0i}) et la pente (b_{1i});
- **(h) Effets cumulés du SARA jusqu'au temps t :** $\kappa_i(b_i, t) = \int_0^{T_i} Y_i(s)ds$ où $Y_i(t) = [Y_i(s), 0 \leq s < t]$; le modèle suppose que le risque de décès au temps t dépend de l'ensemble des mesures du SARA(s) avant t ($0 \leq s < t$).

Ces différents modèles ont été estimés par le package R *JM* et comparés par le critère AIC.

Sous-groupe d'évolution de l'ataxie et leur impact sur la survie

Parce que le taux d'évènement pour les SCA2, SCA3 et SCA6 n'était pas suffisant, nous n'avons pu identifier des sous-groupes d'évolution différente de l'ataxie et de la survie que chez les patients SCA1 en utilisant le modèle conjoint à classe latente [Commenges and Jacqmin-Gadda, 2016, Proust-Lima et al., 2014] décrit dans la section 3.2.2. Contrairement aux modèles conjoints à effets aléatoires partagés, dans le modèle conjoint à classe latente, l'évolution du SARA au cours du temps est différente dans chaque classe latente g pour $g = 1, \dots, G$. Ainsi, la distribution conditionnelle du score SARA en fonction de la classe g pour le patient i à la mesure j est définie par :

$$SARA_{ij} = (X_i^T \beta_{0g} + b_{0i}) + (X_i^T \beta_{1g} + b_{1i}) \times t_{ij} + \epsilon_{ij} \quad (6.3)$$

avec $X = (\text{l'âge à l'inclusion, et le CAG long})$, $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$ et $b_i = (b_{0i}, b_{1i})^T \sim N(0, B_g)$.

Aucun prédicteur n'a été inclut dans la probabilité d'appartenir aux classes, si bien que

$$\pi_{ig} = P(c_i = g) = \frac{e^{\zeta_{0g}}}{\sum_{l=1}^G e^{\zeta_{0l}}} \quad (6.4)$$

avec $\zeta_{0G} = 0$, G étant la classe de référence.

Le risque de décès dans chaque classe g a été décrit par un modèle de survie à risque proportionnel sous l'hypothèse de proportionnalité entre les classes,

$$h_i(t|c_i = g) = h_0(t; \zeta) \exp(W_i^T \alpha_g) \quad (6.5)$$

ajusté sur $W = (\text{âge à l'inclusion et la dysphagie})$ et un risque de base de type Weibull : $h_0(t) = \gamma \lambda^\gamma t^{\gamma-1}$.

Nous avons estimé le modèle conjoint en faisant varier le nombre de classes (1 à 4), puis sélectionné le modèle avec un nombre optimal de classes latente en utilisant un compromis entre le Integrated Likelihood BIC approximation (ICL-BIC plus petit) [Han et al., 2007a] et le non-rejet de l'hypothèse d'indépendance entre la progression de l'ataxie et la survie (p -value du test du score ≥ 0.05) [Jacqmin-Gadda et al., 2010]. Une fois le nombre de classes identifié, nous avons comparé *a posteriori* la distribution des classes par rapport aux variables à l'inclusion afin de déterminer si ces dernières étaient associées aux groupes. En fonction de la nature des variables à l'inclusion, nous avons utilisé soit un test de Chi-deux pour les variables catégorielles soit une analyse de variance pour les variables continues avec le test de Honestly Significant Difference (HSD) pour tenir compte des comparaisons multiples.

Prédiction dynamique de décès

Pour évaluer la capacité du score SARA à prédire le décès, nous avons déduit du modèle conjoint à classe latente (7.5) les prédictions dynamiques individuelles de décès dans une fenêtre $[s, s + t]$ d'horizon de 5 ans ($t = 5$) en utilisant toutes les mesures de SARA collectées $Y_i^{(s)}$ aux temps de prédictions avant 5 ans ($0 \leq s < 5$). Ainsi, pour chaque patient i la probabilité de décès est donnée par la formule :

$$\begin{aligned}
 P(T_i \leq s + t | T_i \geq s, Y_i^{(s)}, H_i; \theta) &= \\
 \sum_{g=1}^G P(T_i \leq s + t | T_i \geq s, c_i = g, H_i; \theta) \times P(c_i = g | T_i \geq s, Y_i^{(s)}, H_i; \theta) &= \\
 = \frac{\sum_{g=1}^G \hat{\pi}_{ig} f(Y_i^{(s)} | c_i = g, H_i; \hat{\theta}) (S_i(s | c_i = g, H_i; \hat{\theta}) - S_i(s + t | c_i = g, H_i; \hat{\theta}))}{\sum_{g=1}^G \hat{\pi}_{ig} f(Y_i^{(s)} | c_i = g, H_i; \hat{\theta}) S_i(s | c_i = g, H_i; \hat{\theta})} & \quad (6.6)
 \end{aligned}$$

où $Y_i^{(s)}$ est le vecteur du score SARA collecté jusqu'à s ; H_i le vecteur de toutes les variables recueillies à l'inclusion ($H_i = (X_i, W_i)^T$), T_i le temps de survie et $\hat{\theta}$ l'estimateur du vecteur de paramètres θ du modèle conjoint.

Nous avons comparé les prédictions dynamiques individuelles dérivées du modèle conjoint avec le nombre optimal de classes latentes à celles estimées à partir du modèle conjoint à 1-classe qui suppose l'indépendance entre la progression de l'ataxie et le risque de décès.

Ensuite, nous avons évalué la capacité prédictive de ces prédictions en utilisant le Expected Prognostic Observed Cross-Entropy (EPOCE) [Commenges et al., 2012] et l'aire sous la courbe dépendant du temps (AUC) avec les intervalles de confiance à 95% de leur différence [Blanche et al., 2015].

Enfin, nous avons fourni un exemple illustratif de la capacité de ces modèles conjoints à faire de la prédiction pour un nouveau patient non inclut dans le modèle conjoint. Ce patient a été sélectionné de manière aléatoire (voire l'article). Toutes les analyses ont été effectuées séparément pour chaque génotype. Les packages R *lcmm* et *timeROC* ont été utilisés pour l'estimation des paramètres.

6.1.3 Résultat

Les résultats de ce travail ont permis de montrer que la structure de dépendance acceptable entre la survie et la progression de l'ataxie était la valeur courante du score SARA quel que soit le génotype. Les modèles conjoints ont permis d'identifier des facteurs de risques communs à la survie et à la progression de la maladie : ce sont l'évolution du SARA (ajustée sur la longueur du CAG pour les SCA1 et l'âge de début de la maladie pour les SCA2), l'âge avancé à l'inclusion et la présence d'une dysphagie pour les SCA1, et l'âge avancé et le nombre de répétitions du CAG long à l'inclusion pour les SCA2. Aucun facteur additionnel n'a été trouvé pour les SCA3 et SCA6.

Pour des raisons de faisabilité, nous n'avons pu faire une recherche de sous-groupes de patients d'évolution différente de l'ataxie et de la survie que pour les patients atteints de SCA1. Trois sous-groupes de patients homogènes en termes d'évolution du SARA et du décès ont été identifiés : groupe de pronostique modéré, intermédiaire et mauvais pronostique. Ce dernier groupe, qui comprend 12% (13/106) des patients était caractérisé par une évolution rapide de l'ataxie avec

un risque de décès multiplié par 11 par rapport au groupe intermédiaire. Selon la classification *a posteriori*, le groupe de mauvais pronostic était plus sévèrement affecté à l'inclusion que les 2 autres. Les patients de ce groupe présentaient plus fréquemment une dysphagie, une dystonie, une déficience cognitive, une altération de leur qualité de vie, un score SARA à l'inclusion et un nombre de répétitions du CAG court plus élevés.

Comparé aux modèles conjoints à effet aléatoire partagé, le modèle conjoint à 3-classes présentait de meilleures performances statistiques soulignant leur flexibilité à modéliser les données longitudinales mais au prix d'une augmentation du nombre de paramètres à estimer. En plus, les prédictions dynamiques individuelles estimées à partir des modèles conjoints à 3-classes, comparées à celles d'1-classe présentaient des estimations plus précises (EPOCE significativement plus bas) et de meilleures capacités prédictives (AUC dépendant du temps plus grandes).

Enfin, nous avons fourni un exemple illustratif d'estimation des prédictions dynamiques individuelles pour un nouveau patient à partir des modèles conjoints à 1 et 3-classes. Globalement, le risque de décès prédit par ces 2 modèles (1 classe et 3 classes) augmente avec l'augmentation du SARA. Mais, les probabilités de décès pour le modèle à 1-classes étaient plus basses.

En résumé, ce travail a permis de quantifier l'impact de la progression de l'ataxie spinocérébelleuse sur la survie et d'identifier les mêmes facteurs de risque pour les plus communs des SCA : les types 1, 2, 3 et 6. En outre, ces résultats ont permis l'identification de sous-groupes de patients SCA1 qui partagent la même progression de la maladie et le même risque de décès. A partir de ces résultats, nous avons développé un outil de prédiction de la survie qui présente de très bonnes capacités prédictives prenant en compte l'ensemble des données collectées du score SARA utilisé pour mesurer la progression de l'ataxie. Cependant, cet outil de prédiction nécessite une validation externe avant d'être utilisés en pratique clinique courante. Enfin, ces résultats pourraient aider à faciliter la sélection et la stratification des patients pour les futurs essais cliniques. L'ensemble de ce travail fera l'objet d'une soumission dans American Journal of Epidemiology. Il est en cours de relecture par les co-auteurs.

6.2 Article 3 :

Prediction of survival with long-term disease progression in spinocerebellar ataxia types 1, 2, 3, and 6

Alhassane Diallo, Heike Jacobi, Arron Cook, Robyn Labrum, Alexandra Durr, Alexis Brice, Perrine Charles, Cecilia Marelli, Caterina Mariotti, Lorenzo Nanetti, Marta Panzeri, Maria Rakowicz, Anna Sobanska, Anna Sulek, Tanja Schmitz-Hübsch, Ludger Schöls, Holger Hengel, Bela Melegh, Alessandro Filla, Antonella Antenora, Jon Infante, José Berciano, Bart P. van de Warrenburg, Dagmar Timmann, Sylvia Boesch, Massimo Pandolfo, Jörg B. Schulz, Peter Bauer, Paola Giunti, Kang Jun-Suk, Laszlo Baliko, Michael H. Parkinson, Thomas Klockgether, and Sophie Tezenas du Montcel

Correspondance to Dr. Sophie Tezenas du Montcel, INSERM, Sorbonne Universités, Assistance Publique – Hôpitaux de Paris AP-HP, Hôpitaux Universitaires Pitié-Salpêtrière – Charles Foix, F-75013, Paris, France

sophie.tezenas@aphp.fr; Phone: +33142160582.

Authors affiliations: INSERM, Sorbonne Universités, Institut Pierre Louis d'Epidémiologie et de Santé Publique, IPLESP, F-75013, Paris, France (Alhassane Diallo) ; Department of Neurology, University Hospital of Heidelberg, Heidelberg, and German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany (Heike Jacobi, Tanja Schmitz-Hübsch, Thomas Klockgether) ; Department of Molecular Neuroscience, UCL, Institute of Neurology, London, United Kingdom (Arron Cook, Paola Giunti, Michael H. Parkinson); Neurogenetics Laboratory, National Hospital of Neurology and Neurosurgery, UCLH, London, United Kingdom (Robyn Labrum); Sorbonne Universités, UPMC Univ Paris 06, UMRS_1127, INSERM, U 1127, CNRS, UMR 7225, Institut du Cerveau et de la Moelle épinière, ICM, Pitié-Salpêtrière University Hospital Paris, France (Alexandra Durr, Alexis Brice) ; APHP, Genetics Department, Pitié-Salpêtrière University Hospital Paris, France (Alexandra Durr, Alexis Brice) ; Service de Neurologie – CMRR, CHRU Gui de Chauliac, 80, av. A. Fliche, 34295 - Montpellier CEDEX 05, France (Cecilia Marelli) ; SOSD Genetics of Neurodegenerative and Metabolic Diseases, Fondazione-IRCCS Istituto Neurologico Carlo Besta, Milan, Italy (Caterina Mariotti, Lorenzo Nanetti, Marta Panzeri) ; Department of Clinical Neurophysiology, Institute of Psychiatry and Neurology, Warsaw, Poland (Maria Rakowicz, Anna Sobanska) ; Department of Genetics, Institute of Psychiatry and Neurology, Warsaw, Poland (Anna Sulek) ; Charité-Universitätsmedizin Berlin, NeuroCure Clinical Research Center, Clinical Neuroimmunology Group, Berlin, Germany (Tanja Schmitz-Hübsch) ; Department of Neurodegeneration and Hertie-Institute for Clinical Brain Research, University of Tübingen and Deutsches

Zentrum für Neurodegenerative Erkrankungen (DZNE), D-72076 Tübingen, Germany (Ludger Schöls, Holger Hengel) ; Department of Medical Genetics, and Szentagothai Research Center, University of Pécs, Pécs, Hungary (Bela Melegh) ; Department of Neuroscience, and Reproductive and Odontostomatological Sciences, Federico II University Naples, Italy (Alessandro Filla, Antonella Antenora) ; Service of Neurology, University Hospital Marqués de Valdecilla (IDIVAL), University of Cantabria (UC) and Centro de Investigación Biomédica en Red de Enfermedades Neurodegenerativas (CIBERNED), Santander, Spain (Jon Infante, José Berciano) ; Department of Neurology, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, PO Box 9101, 6500 HB Nijmegen, The Netherlands (Bart P. van de Warrenburg) ; Department of Neurology, Essen University Hospital, University of Duisburg-Essen (Dagmar Timmann) ; Department of Neurology, Medical University, Innsbruck, Innsbruck Austria (Sylvia Boesch) ; Université Libre de Bruxelles (ULB), Neurology Service - ULB Hôpital Erasme, ULB Laboratory of Experimental Neurology, Brussels, Belgium (Massimo Pandolfo) ; Department of Neurology, RWTH Aachen University, Pauwelsstraße 30, 52074 Aachen, Germany and JARA - Translational Brain Medicine, Aachen-Jülich, INM 11, Germany (Jörg B. Schulz) ; Department of Neurology, Zala County Hospital, H-8900 Zalaegerszeg, Zrinyi M. Str. 1, Hungary (Bela Melegh, Laszlo Baliko) ; Department of Neurology, University of Frankfurt, Frankfurt/M, Germany (Kang Jun-Suk) ; Department of Neurology, University Hospital of Bonn, Bonn, Germany and German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany (Thomas Klockgether, Ludger Schöls, Holger Hengel) ; INSERM, Sorbonne Universités, Assistance Publique – Hôpitaux de Paris AP-HP, Hôpitaux Universitaires Pitié-Salpêtrière – Charles Foix, F-75013, Paris, France (Sophie Tezenas du Montcel).

Funding: EU FP6 (EUROSCA, grant LSHM-CT-503304), German Ministry of Education and Research (GeneMove), Polish Ministry of Scientific Research and Information Technology, EU FP7 (Neuromics, grant F5-2012-305121), and Fondation pour la Recherche Médicale (FRM, grant number: PLP20151034334).

Conflicts of interests: STdM reports grants from EU FP6 EUROSCA, during the conduct of the study; ADi reports grants from Fondation pour la Recherche Médicale (FRM), grant number: PLP20151034334, during the conduct of the study; TK reports grants from EU FP6 EUROSCA, during the conduct of the study; MP reports grants from Friedreich Ataxia Research Alliance, grants from FNRS (Belgium), grants and personal fees from Biomarin, grants and personal fees from Voyager Therapeutics, personal fees from Apopharma, personal fees from Vertex, personal fees from Pfizer, grants from Euroataxia, outside the submitted work; BPvdW reports grants from Radboud university medical centre, grants from Hersenstichting, grants from ZonMW, grants from

BBMRI-NL, grants from Gossweiler Foundation, grants from Bioblast Pharma, outside the submitted work; PB reports grants from European Commission, during the conduct of the study; personal fees from CENTOGENE AG, personal fees from Actelion Pharmaceuticals, outside the submitted work; TS-H reports grants from EU, during the conduct of the study; grants from Bundesministerium für Wirtschaft und Energie, grants from Ipsen Pharma, outside the submitted work; JBS serves on scientific advisory boards for Lundbeck Inc., TEVA, Novartis, ForwardPharma and Lilly; he has received funding for travel and speaker honoraria from GlaxoSmithKline, Merz Pharmaceuticals, Medical Tribune, Lundbeck Inc, Pfizer Inc, Boehringer, Bayer; serves as Editor-in-Chief of the Journal of Neurochemistry and Associate Editor for eNeuro; PG is supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre; DT was funded by the German Research Foundation (DFB), EU, the Bernd Fink Stiftung and the German Heredoataxia-Foundation (DHAG) during the time of the research (all funding unrelated to the study), and JB, AA, AF, HH, LS, RL, MR, AC, ASu, ASo, HJ, SB, JI, CaM, LN, MarP, BM, ADu, AB, CeM, PC, JB, LB, MHP and K-JS declare no competing interests.

Running head: Rapid Ataxia Progression is Related to Poor Survival

Abstract

Spinocerebellar ataxias (SCA) are rare dominantly inherited neurodegenerative diseases that lead to severe disability and premature death. We aimed to quantify the impact of the disease progression on survival and to identify different profiles of disease progression and survival. From EUROSCA longitudinal cohort study, participants with progressive, otherwise unexplained ataxia, and who had positive genetic test for SCA1, SCA2, SCA3, or SCA6 were enrolled. Outcomes were the change over time in total SARA score, and time-to-death. Joint model was used to analyze disease progression and survival. In all genotype, disease progression was the strongest predictor for death: an increase of one standard deviation (SD) in total SARA score increased the risk of death by 1.28 times (95% CI: 1.18, 1.38) for patients with SCA1; 1.19 times (1.12, 1.26) for SCA2; 1.30 times (1.19, 1.42) for SCA3; and 1.26 times (1.11, 1.43) for SCA6. Three subgroups of disease progression and survival were identified for patients with SCA1: “severe” (n=13, 12%), “intermediate” (n=31, 29%) and “moderate” (n=62, 58%). Severe group had the worse prognosis at baseline. Rapid ataxia progression is associated with poor survival of the most common SCA. These results have implication for the design of future interventional studies of SCA.

Keywords: EUROSCA study, spinocerebellar ataxia, longitudinal data, join model, dynamic predictions.

Abbreviations: SCA, Spinocerebellar Ataxia, CAG, Cysteine Alanine Glycine, SARA, Scale for the Assessment and Rating of Ataxia, INAS, Inventory of Non-Ataxia Signs, JLCM, Joint Latent Class Model, SREM, Shared Random-Effects Model, EPOCE, Expected Prognostic Observed Cross-Entropy, AUC, Area Under Curve.

Spinocerebellar ataxias (SCAs) are inherited rare neurological disorders, characterized clinically by progressive balance problems and incoordination that lead to severe disability and premature death. They are caused by abnormal CAG repeat expansions, encoding elongated polyglutamine tracts within the proteins associated with each type. More than 40 genetically distinct SCAs have been defined with the most common being SCA1, SCA2, SCA3 and SCA6 (1).

The European prospective cohort study (EUROSCA) of patients with SCA1, SCA2, SCA3 and SCA6 was initiated in 2005 to study the natural history of ataxia. Longitudinal clinical (2) and survival data (3) from EUROSCA allowed us to establish genotype-specific progression and survival rates, and identify factors that influence disease course and survival. These analyses identified SCA1 as the genotype with the fastest disease progression and shortest survival among these SCAs (2, 3). Simultaneously modeling long-term disease progression and survival would allow taking into account their correlation, better understanding of their dependence, identification of common prognostic factors, and could advance design of future clinical trials.

Here, we report data from EUROSCA study with long-term follow-up on disease progression in relation to survival. We aimed to (a) quantify the effect of the disease progression on survival of patients with SCA1, SCA2, SCA3 or SCA6, (b) identify different subgroups of disease progression and survival as well as factors influencing these groups, and (c) compute and compare the individual dynamic prediction of survival based on disease progression.

METHODS

Participants

From the 525 SCAs patients enrolled between July 1, 2005 and Aug 31, 2006 in the multicenter (17 European centers) EUROSCA prospective cohort study (4), a subsample of 462 SCAs patients with at least two visits were analysed. These patients were diagnosed with progressive, otherwise unexplained ataxia, and a positive molecular genetic test for SCA1 (n=107), SCA2 (n=146), SCA3 (n=122) or SCA6 (n=87). Assessments were performed according to a written study protocol. Patients were seen at baseline and followed by annual visits for three years. Afterwards, study participants entered an extension phase in which study assessments were performed during routine visits, resulting in irregular intervals between the visits. The database was locked in November 03, 2016, after a maximum observation period of 11 years. The study was approved by the ethics committees of the participating centres. Written informed consent was obtained from all study participants at enrollment.

Outcomes Measures

The clinical outcomes were disease progression and overall survival. As the recent previous survival study (3), we updated the vital status to avoid survival bias from available electronic EUROSCA patient registry or by interviews of family members or by questioning of records from civil registry offices when feasible. Disease progression was measured by the Scale for the Assessment and Rating of Ataxia (SARA) (4). Briefly, the SARA score consists in eight items that yield a total score of 0 (no ataxia) to 40 (most severe ataxia). All investigators were experienced in the use of SARA.

Covariates

As candidate predictors to joint disease progression and survival modeling, we selected age at baseline, gender, age at ataxia onset, repeat lengths of the expanded alleles, SARA score at baseline, dysphagia and dystonia, which have been reported as predictors for survival or disease progression in our previously published studies (2, 3). As additional candidates, we selected the Inventory of Non-Ataxia Signs (INAS) (5) and its reported abnormalities such as double vision and cognitive impairment, a physical state (body mass index (BMI)), and the health-related quality of life assessed by the visual analogical scale (VAS) of the EuroQol five dimensions questionnaire (EQ-5D) (6, 7). BMI was calculated using the formula ($\text{weight}/\text{height}^2$). Repeat lengths of the expanded and normal alleles determined at the Institute of Medical Genetics and Applied Genomics of the University of Tübingen (Tübingen, Germany).

Statistical Analysis

Frequencies (percentages) or mean (standard deviation) were used to describe the categorical and continuous variables at baseline. Time from enrollment was used as the time scale. Survival was calculated from the date of enrollment to death for any reason. Data for patients who were alive or lost to follow-up were censored. We simultaneously modeled time-to-death and disease progression, using the Shared Random-Effects (SRE) model (8) to quantify and capture the best relationship between disease progression and survival, and the Joint Latent Class (JLC) model (9) to identify subgroups of survival and disease progression. In each joint model, the survival model was the Weibull proportional regression model adjusted on age at inclusion, and the longitudinal model was a linear mixed-effects regression model. From the SRE model, we explored several dependence structures (appendix, p27) in order to capture the best relationship between disease progression and survival. Thereafter, we selected the strongest contributing predictors for survival from a multivariate regression through a backward procedure based on the lowest Akaike Information Criterion.

Because the proportion of death was small in SCA2, SCA3 and SCA6 compared to SCA1 patients, we did a search for subgroups of participants sharing the same profiles of disease progression and the risk of death using the JLC model only for patients with SCA1. To select the best-fitting model with the optimal number of latent classes, we used a compromise between the lower Integrated Classification Likelihood with BIC approximation (ICL-BIC) (10) and the non-significant score test for the conditional independence assumption between longitudinal and survival outcomes (11). Distributions of the baseline variables across these classes were compared a posteriori using a chi-squared (X^2) test for the categorical variables and an analysis of variance for continuous variables with the Honestly Significant Difference test to account for multiple comparisons.

To assess the ability of the longitudinal SARA score to predict death, we derived from the best-fitting JLC model and the 1-class JLC model which assuming independence between SARA and the risk of death, the individual dynamic predictions of death. These predictions are updated at each new measurement of SARA in a time horizon of 5-year using all measures of SARA collected till time at prediction s (before 5-year). The choice of a 5-year prediction horizon was guided by the feasibility and the clinical perspective of targeting patients at high risk of disease progression and death, who could benefit from future treatment. To compare the predictive accuracy and the discriminative ability of these predictions, we used the Expected Prognostic Observed Cross-Entropy (EPOCE) (12) and the time-dependent Area under curve (AUC) (13) as well as the 95% confidence intervals of the difference in EPOCE and AUC. For EPOCE measure, the lower the better, and conversely for AUC the higher the better.

We provided an example for a new patient, randomly selected and not included to build the model, to illustrate how the joint modeling can be used in practice to do individual dynamic predictions. Subsequent analyses were performed separately for each genotype, as survival and disease progression differed between genotypes (2, 3). All data analyses were performed using the “*JM*” and “*lcmm*” R packages. Values of $P < 0.05$ were considered to be statistically significant and all tests were two-sided.

RESULTS

Population Characteristics

The 462 patients analyzed in the current study had a total of 2192 visits, with a median of 3 visits (interquartile range, 2 – 4 visits) for each patient. Individual profiles of the SARA evolution show that deceased patients have on average a higher SARA score at baseline (appendix, p25). Demographic and clinical data at baseline are provided in Table 1. During the follow-up, 102 (22%) patients died: 32 (30%) with SCA1, 33 (23%) with SCA2, 24 (21%) with SCA3, and 11 (13%) with SCA6.

Disease Progression and Survival Relationship

To explore the best link between disease progression and survival, we applied the shared random-effect (SRE) model. For all genotypes, we found that the current value of the SARA score at a particular visit t was the best relationship with the risk of death at the same visit point according to the AIC (appendix p27). In SCA1 patients, a predictive model obtained from the multivariate joint model identified age at baseline, dysphagia and the current value of the SARA score as the strongest contributing risk factors for death (Table 2). An increase of one standard deviation (SD) in the total SARA score was associated with a 28% increase in risk of death. For SCA2, the corresponding risk factors for death were age at baseline, longer CAG repeat number, and the current value of the SARA score. The risk of death increased by 19% for one SD increase in the SARA score (Table 2). For SCA3 and SCA6, the only contributing risk factor for death was the current value of the SARA score. An increase of one SD in the SARA score increased the risk of death by 30% for SCA3 patients, and 26% for SCA6 respectively.

Sub-groups of Disease Progression and Survival

To identify subgroups of patients sharing the same profiles of disease progression and the risk of death, the JLC model with several latent classes ranging from 1 to 4 classes were estimated for patients with SCA1. The model with the optimal number of classes selected by the compromise criterion included three classes (appendix, p28). A class-specific trajectories and survival function displays in Figure 1 (A and B) showed a small group (severe group) representing 12% (n=13) of patients with faster progression of SARA score (4.67 [SE 2.96]) associated with a high risk of death relative to the intermediate group (HR: 11.28, 95% CI: 3.75, 33.93; $P < .0001$). The other groups representing 29% and 59% of the patients corresponded to different evolutions of SARA associated with risks of death from intermediate to moderate. According to the posterior classification provided in Table 3, the patients in the severe group were significantly more affected at baseline compared to the intermediate and

moderate groups. Thus, the severe group was characterized by a lower health-related quality of life, body mass index (BMI), and CAG repeat length of the normal allele, and a longer disease duration, a higher SARA score and frequency of number of non-ataxia signs, and the following individual non-ataxia signs: dystonia, cognitive impairment and dysphagia.

Dynamic Predictions

We derived from the 1-class and 3-class JLC model, the individual dynamic prediction of survival at different times of prediction s (before 5-year) in the window $[s, s+5]$ years, and compared them (Figure 2). According to the 95% confidence intervals of the difference in EPOCE and AUC (excluded zero), we found that the 3-class JLC model had a more accurate estimate (lower EPOCE) and a better discriminative ability to detect patients with high risk of death (higher time-dependent AUC). This ability of the SARA score to discriminate subject with high-risk of death exceeded 85% (AUC 0.852 [SD 0.062], 0.965 [0.026], 0.974 [0.016], and 0.952 [0.025] for visit-1, visit-2, visit-3 and visit-4 prediction times respectively). To illustrate the individual dynamic prediction of survival from the joint model, we considered a virtual patient who entered the cohort at 27 years old and died at 32 years. Overall, the risk of death predicted by the JLC model (1-class and 3-class) increased with the increasing SARA score trajectories, while the predictions from the 1-class JLC model remain very low (appendix, p26).

DISCUSSION

To our knowledge, this is the first longitudinal study that addresses the joint disease progression and survival modeling among patients with spinocerebellar ataxia. With this approach, we were able to identify of subgroups with respect disease progression and survival for patients with SCA1, and to develop an individual dynamic prediction tool of survival.

We assessed the progression of ataxia using the SARA score, which was validated in a rigorous process (4), and is more sensitive to change compared to the number of non-ataxia signs (INAS) (14). Moreover, the progression of ataxia was strongly associated with death in all disease subtype underlining the clinical relevance and predictive power of SARA. Although, the INAS score allows capturing additional information on neurological symptoms other than ataxia, we did not find a significant effect of the INAS score progression on survival for any genotype (data not shown).

To quantify the effect of disease progression on survival, we tested several dependences structures between two outcomes. In all genotypes, we found that survival was more influenced by the severity of ataxia at a given time point t (current value) than the by the rate of disease progression. The strongest contributing risk factors of death from the joint model were similar to those previously identified by Cox model except for SCA3 patients (3). We failed to replicate the significant negative interaction between age at baseline and the CAG repeat number probably due to a lack of power following the restricted analysis to patients who had at least two visits, which lead to a decrease from 34 to 26 the number of events. For SCA2 patients, in addition to the current value of the SARA score, the risk of death increases with longer CAG repeat number and older age at inclusion. Similar results were obtained in a recent retrospective analysis of Italian SCA2 patients (15).

Compared to the SRE model, the JLC model gave a substantial gain in goodness-of-fit. The maximal difference in log-likelihood between the SRE model with current value and the 3-class JLC model was 20.4. This difference illustrates the flexibility of the JLC model to model the disease progression, the time-to-death and their dependence. However, this flexibility was associated with a large increase of the number of parameters: 24 for the 3-class JLC model vs. 14 for the SRE model, suggesting some over-fitting (appendix, p28). In addition, unlike the SRE model, which quantifies an average impact of the disease progression on survival, the JLC model identified three profiles of the SCA1 patients. The average probability to belong to the 3-class JLC model was higher, ranging from 0.89 to 0.98, suggesting an unambiguous classification (appendix, p29). Patients from these groups share the same ataxia evolution and risk of death with 12% of the subjects attributed to the severe group.

The most important finding was that the severe group had faster disease progression and higher risk of death. Moreover, the patients from the severe group have a lower CAG repeat number of the normal allele highlighting the modifier effect of the normal *ATXN1* allele in SCA1. A deleterious effect of shorter length of the normal allele in SCA1 has been previously shown for age at onset (16), but not for disease progression (2). These patients also had a lower body weight at baseline compared to the other groups, which corroborates the recent reported data on weight loss in relation to disease progression. Investigators showed that patients who lose weight have faster disease progression (17).

Furthermore, we derived a subject-specific dynamic prediction of survival from the JLC models (1-class and 3-class), and we showed that the 3-class JLC model provides a more precise estimate and a good predictive accuracy compared to the 1-class JLC model which assumes independence between SARA and risk of death. These findings highlight the importance of subgroup analyzes that release the strong hypothesis of homogeneity of the population assumed by the 1-class JLC model, and confirms the presence of a clinical heterogeneity and different profiles of ataxia progression within the same genotype. Considering this heterogeneity will help to improve the design of future clinical trials.

In our published survival data from the same cohort, we developed a prognostic nomogram score from the Cox model. On the basis of the prognostic nomogram, we grouped it into three risk categories: good, intermediate or poor prognosis (3). To assess the agreement between the 3-class JLC and Cox models in term of classification of SCA1 patients into three groups, we computed the kappa coefficient with 95% tracking confidence interval assuming the joint model as the reference. Overall, we found a good rate of agreement between the two approaches (74% [95% CI: 64, 85]). Disagreement concerned 13 patients in the good prognosis group from the Cox model, attributed to the intermediate group by the 3-class JLC model or vice versa, and five patients in the intermediate prognosis group from the 3-class JLC model, classified as having a poor prognosis by the Cox model. However, the agreement for the classification of the poor prognosis group between the two approaches was perfect. Although both approaches show an acceptable agreement rate, the joint model has several advantages. First, unlike the Cox model, the joint model allows taking into account the repeated measurement of the clinical scales (SARA) by correcting the bias induced by the random measurement error and the occurrence of the event (death) (18). Second, the joint model provides a better understanding of the link between the repeated clinical scores and the risk of event. Third, it allows the development of powerful dynamic prognostic tools (19, 20).

Strengths of our report include the large number of patients and the long-term observational period up to 10 years. Despite an average follow-up of 20 years after ataxia onset, this 10 years observation period might seem short in view of an estimated survival of patients with spinocerebellar ataxia of 20-30 years after onset (21, 22). As in our previous survival analysis (3), the information about vital status was incomplete. To reduce the bias due to censorship, we updated the vital status via the records from civil registry offices and the interview of family members, which allowed us to recover 31% (32/102) of the death. Another strength of our study was the simultaneous disease progression and survival modeling using the joint model methodology, which was robust against substantial dropout rate previously reported, mainly for disease-related reason during the open extension period (2). However, the restriction to clinical assessment, the absence of long-term biomarkers and imaging data, and the scarcity of death events negating the possibility of performing a subgroup analysis for SCA2, SCA3 and SCA6 were the limits of this study.

These data extend the knowledge of the biological characteristics of patients with SCA1, SCA2, SCA3, and SCA6. Furthermore, the results provide more information to facilitate selection and stratification of patients for future interventional clinical trials.

ACKNOWLEDGMENTS

Authors affiliations: INSERM, Sorbonne Universités, Institut Pierre Louis d'Epidémiologie et de Santé Publique, IPLESP, F-75013, Paris, France (Alhassane Diallo) ; Department of Neurology, University Hospital of Heidelberg, Heidelberg, and German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany (Heike Jacobi, Tanja Schmitz-Hübsch, Thomas Klockgether) ; Department of Molecular Neuroscience, UCL, Institute of Neurology, London, United Kingdom (Arron Cook, Paola Giunti, Michael H. Parkinson); Neurogenetics Laboratory, National Hospital of Neurology and Neurosurgery, UCLH, London, United Kingdom (Robyn Labrum); Sorbonne Universités, UPMC Univ Paris 06, UMRS_1127, INSERM, U 1127, CNRS, UMR 7225, Institut du Cerveau et de la Moelle épinière, ICM, Pitié-Salpêtrière University Hospital Paris, France (Alexandra Durr, Alexis Brice) ; APHP, Genetics Department, Pitié-Salpêtrière University Hospital Paris, France (Alexandra Durr, Alexis Brice) ; Service de Neurologie – CMRR, CHRU Gui de Chauliac, 80, av. A. Fliche, 34295 - Montpellier CEDEX 05, France (Cecilia Marelli) ; SOSD Genetics of Neurodegenerative and Metabolic Diseases, Fondazione-IRCCS Istituto Neurologico Carlo Besta, Milan, Italy (Caterina Mariotti, Lorenzo Nanetti, Marta Panzeri) ; Department of Clinical Neurophysiology, Institute of Psychiatry and Neurology, Warsaw, Poland (Maria Rakowicz, Anna Sobanska) ; Department of Genetics, Institute of Psychiatry and Neurology, Warsaw, Poland (Anna Sulek) ; Charité-Universitätsmedizin Berlin, NeuroCure Clinical Research Center, Clinical Neuroimmunology Group, Berlin, Germany (Tanja Schmitz-Hübsch) ; Department of Neurodegeneration and Hertie-Institute for Clinical Brain Research, University of Tübingen and Deutsches Zentrum für Neurodegenerative Erkrankungen (DZNE), D-72076 Tübingen, Germany (Ludger Schöls, Holger Hengel) ; Department of Medical Genetics, and Szentagothai Research Center, University of Pécs, Pécs, Hungary (Bela Melegh) ; Department of Neuroscience, and Reproductive and Odontostomatological Sciences, Federico II University Naples, Italy (Alessandro Filla, Antonella Antenora) ; Service of Neurology, University Hospital Marqués de Valdecilla (IDIVAL), University of Cantabria (UC) and Centro de Investigación Biomédica en Red de Enfermedades Neurodegenerativas (CIBERNED), Santander, Spain (Jon Infante, José Berciano) ; Department of Neurology, Donders Institute for Brain, Cognition and Behaviour, Radboud University Medical Center, PO Box 9101, 6500 HB Nijmegen, The Netherlands (Bart P. van de Warrenburg) ; Department of Neurology, Essen University Hospital, University of Duisburg-Essen (Dagmar Timmann) ; Department of Neurology, Medical University, Innsbruck, Innsbruck Austria (Sylvia Boesch) ; Université Libre de Bruxelles (ULB), Neurology Service - ULB Hôpital Erasme, ULB Laboratory of Experimental Neurology, Brussels, Belgium (Massimo Pandolfo) ; Department of Neurology, RWTH Aachen University, Pauwelsstraße

30, 52074 Aachen, Germany and JARA - Translational Brain Medicine, Aachen-Jülich, INM 11, Germany (Jörg B. Schulz) ; Department of Neurology, Zala County Hospital, H-8900 Zalaegerszeg, Zrinyi M. Str. 1, Hungary (Bela Melegh, Laszlo Baliko) ; Department of Neurology, University of Frankfurt, Frankfurt/M, Germany (Kang Jun-Suk) ; Department of Neurology, University Hospital of Bonn, Bonn, Germany and German Center for Neurodegenerative Diseases (DZNE), Bonn, Germany (Thomas Klockgether, Ludger Schöls, Holger Hengel) ; INSERM, Sorbonne Universités, Assistance Publique – Hôpitaux de Paris AP-HP, Hôpitaux Universitaires Pitié-Salpêtrière – Charles Foix, F-75013, Paris, France (Sophie Tezenas du Montcel).

Contributors: ADi designed and executed the statistical analysis, and wrote the first draft of the report and reviewed the report; HJ and TS-H contributed to the conception, organisation, and execution of the research project and reviewed and commented on the statistical analysis and the report; AC, RL, ADu, AB, PC, CeM, CaM, LN, MarP, MR, ASo, ASu, LS, HH, BM, AF, AA, JI, JB, BPvdW, DT, SB, MP, JBS, PB, PG, LB, MHP and K-JS organised and did the research project and reviewed and commented on the statistical analysis and the report; TK conceived, organized, and did the research project, and designed, reviewed, and commented on the statistical analysis and wrote the first draft of the report and reviewed the report; STdM contributed to the conception of the research project and wrote the first draft of the report and reviewed the report.

Funding: EU FP6 (EUROSCA, grant LSHM-CT-503304), German Ministry of Education and Research (GeneMove), Polish Ministry of Scientific Research and Information Technology, EU FP7 (Neuromics, grant F5-2012-305121), and Fondation pour la Recherche Médicale (FRM, grant number: PLP20151034334).

Thanks: This study was funded within the framework of EU FP6 (EUROSCA). Additional funds came from the German Ministry of Education and Research (BMBF; GeneMove), Polish Ministry of Scientific Research and Information Technology, and from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement number 2012-305121 "Integrated European –omics research project for diagnosis and therapy in rare neuromuscular and neurodegenerative diseases (NEUROMICS)". We thank Thomas Klopstock (Department of Neurology, University of Munich, Munich, Germany) and Jens Petersen (Department of Neurology, University of Munich, Munich, Germany, Department of Neurology, University of Zurich, Zurich, Switzerland), and Catherine Delnooz (Department of Neurology, Radboud University Nijmegen Medical Center, Nijmegen, Netherlands) and Rafal Rola (First Department of Neurology, Institute of Psychiatry and Neurology,

Warsaw, Poland) for contribution of patients and help in patient assessment. We thank the Fondation pour la Recherche Médicale for the financial support of Alhassane Diallo, grant number: PLP20151034334.

Conflicts of interests: STdM reports grants from EU FP6 EUROSCA, during the conduct of the study; ADi reports grants from Fondation pour la Recherche Médicale (FRM), grant number: PLP20151034334, during the conduct of the study; TK reports grants from EU FP6 EUROSCA, during the conduct of the study; MP reports grants from Friedreich Ataxia Research Alliance, grants from FNRS (Belgium), grants and personal fees from Biomarin, grants and personal fees from Voyager Therapeutics, personal fees from Apopharma, personal fees from Vertex, personal fees from Pfizer, grants from Euroataxia, outside the submitted work; BPvdW reports grants from Radboud university medical centre, grants from Hersenstichting, grants from ZonMW, grants from BBMRI-NL, grants from Gossweiler Foundation, grants from Bioblast Pharma, outside the submitted work; PB reports grants from European Commission, during the conduct of the study; personal fees from CENTOGENE AG, personal fees from Actelion Pharmaceuticals, outside the submitted work; TS-H reports grants from EU, during the conduct of the study; grants from Bundesministerium für Wirtschaft und Energie, grants from Ipsen Pharma, outside the submitted work; JBS serves on scientific advisory boards for Lundbeck Inc., TEVA, Novartis, ForwardPharma and Lilly; he has received funding for travel and speaker honoraria from GlaxoSmithKline, Merz Pharmaceuticals, Medical Tribune, Lundbeck Inc, Pfizer Inc, Boehringer, Bayer; serves as Editor-in-Chief of the Journal of Neurochemistry and Associate Editor for eNeuro; PG is supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre; DT was funded by the German Research Foundation (DFB), EU, the Bernd Fink Stiftung and the German Heredoataxia-Foundation (DHAG) during the time of the research (all funding unrelated to the study), and JB, AA, AF, HH, LS, RL, MR, AC, ASu, ASo, HJ, SB, JI, CaM, LN, MarP, BM, ADu, AB, CeM, PC, JB, LB, MHP and K-JS declare no competing interests.

REFERENCES:

1. Durr A. Autosomal dominant cerebellar ataxias: polyglutamine expansions and beyond. *Lancet Neurol.* 2010;9(9):885-94.
2. Jacobi H, du Montcel ST, Bauer P, Giunti P, Cook A, Labrum R, et al. Long-term disease progression in spinocerebellar ataxia types 1, 2, 3, and 6: a longitudinal cohort study. *Lancet Neurol.* 2015;14(11):1101-8.
3. Diallo A, Jacobi H, Cook A, Labrum R, Durr A, Brice A, et al. Survival in patients with spinocerebellar ataxia types 1, 2, 3, and 6 (EUROSCA): a longitudinal cohort study. *Lancet Neurol.* 2018;17(4):327-34.
4. Schmitz-Hubsch T, du Montcel ST, Baliko L, Berciano J, Boesch S, Depondt C, et al. Scale for the assessment and rating of ataxia: Development of a new clinical scale. *Neurology.* 2006;66(11):1717-20.
5. Schmitz-Hübsch T, Coudert M, Bauer P, Giunti P, Globas C, Baliko L, et al. Spinocerebellar ataxia types 1, 2, 3, and 6: disease severity and nonataxia symptoms. *Neurology.* 2008;71(13):982-9.
6. EuroQol Group. EuroQol--a new facility for the measurement of health-related quality of life. *Health Policy Amst Neth.* 1990;16(3):199-208.
7. Whynes DK. Correspondence between EQ-5D health state classifications and EQ VAS scores. *Health Qual Life Outcomes.* 2008;6(1):94.
8. Rizopoulos D. Joint models for longitudinal and time-to-event data: with applications in R. Boca Raton: *CRC Press*; 2012.
9. Proust-Lima C, Séne M, Taylor JMG, Jacqmin-Gadda H. Joint latent class models for longitudinal and time-to-event data: a review. *Stat Methods Med Res.* 2014;23(1):74-90.
10. Han J, Slate EH, Peña EA. Parametric latent class joint model for a longitudinal biomarker and recurrent events. *Stat Med.* 2007;26(29):5285-302.
11. Jacqmin-Gadda H, Proust-Lima C, Taylor JMG, Commenges D. Score Test for Conditional Independence Between Longitudinal Outcome and Time to Event Given the Classes in the Joint Latent Class Model. *Biometrics.* 2010;66(1):11-9.
12. Commenges D, Liqueur B, Proust-Lima C. Choice of Prognostic Estimators in Joint Models by Estimating Differences of Expected Conditional Kullback-Leibler Risks. *Biometrics.* 2012;68(2):380-7.
13. Blanche P, Proust-Lima C, Loubère L, Berr C, Dartigues J-F, Jacqmin-Gadda H. Quantifying and comparing dynamic predictive accuracy of joint models for longitudinal marker and time-to-event in presence of censoring and competing risks: Comparing Dynamic Predictive Accuracy of Joint Models. *Biometrics.* 2015;71(1):102-13.

14. Schmitz-Hubsch T, Fimmers R, Rakowicz M, Rola R, Zdzienicka E, Fancellu R, et al. Responsiveness of different rating instruments in spinocerebellar ataxia patients. *Neurology*. 2010;74(8):678-84.
15. Antenora A, Bruzzese D, Lieto M, Roca A, Florio MT, Peluso S, et al. Predictors of survival in spinocerebellar ataxia type 2 population from Southern Italy. *Neurol Sci*. 2018.
16. Tezenas du Montcel S, Durr A, Bauer P, Figueroa KP, Ichikawa Y, Brussino A, et al. Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes. *Brain*. 2014;137(9):2444-55.
17. Diallo A, Jacobi H, Schmitz-Hübsch T, Cook A, Labrum R, Durr A, et al. Body Mass Index Decline Is Related to Spinocerebellar Ataxia Disease Progression. *Mov Disord Clin Pract*. 2017;4(5):689-97.
18. Prentice RL. Covariate Measurement Errors and Parameter Estimation in a Failure Time Regression Model. *Biometrika*. 1982;69(2):331.
19. Proust-Lima C, Taylor JMG. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA: a joint modeling approach. *Biostatistics*. 2009;10(3):535-49.
20. Taylor JMG, Yu M, Sandler HM. Individualized Predictions of Disease Progression Following Radiation Therapy for Prostate Cancer. *J Clin Oncol*. 2005;23(4):816-25.
21. Monin M-L, Tezenas du Montcel S, Marelli C, Cazeneuve C, Charles P, Tallaksen C, et al. Survival and severity in dominant cerebellar ataxias. *Ann Clin Transl Neurol*. 2015;2(2):202-7.
22. Klockgether T, Lütke R, Kramer B, Abele M, Bürk K, Schöls L, et al. The natural history of degenerative ataxia: a retrospective study in 466 patients. *Brain J Neurol*. 1998;121 (Pt 4):589-600.

Figure 1: Class-specific mean trajectories and class-specific predicted survival probabilities from the 3-class JLC model for longitudinal SARA score in SCA1 patients. (A): Weighted subject-specific predicted for SARA trajectories according to the joint latent class model equation: severe (dotted line) = $18.5 + 4.7 * \text{time} + 0.27 * \text{CAG} - 0.07 * \text{interaction between time and repeat CAG}$; intermediate (dashed line) = $26.4 - 0.94 * \text{time} - 0.16 * \text{CAG} + 0.07 * \text{interaction between time and repeat CAG}$; moderate (solid line) = $9.7 - 1.86 * \text{time} - 0.006 * \text{CAG} + 0.08 * \text{interaction between time and repeat CAG}$; (B): Predicted survival according to the Weibull proportional hazard model equation with the intermediate group as reference (dashed line): severe (dotted line) = $0.27530^2 * 1.76747^2 * ((0.27530^2 * \text{time})^{(1.76747^{2-1})}) * e^{2.42274}$; stable (solid line) = $0.27530^2 * 1.76747^2 * ((0.27530^2 * \text{time})^{(1.76747^{2-1})}) * e^{-3.43041}$;

Figure 2: Comparison of the Predictive Accuracy of the two-predicted risk of death within time window (s, s+1) when s = [visits before 5-year] and t=5 years from JLCM estimated on SCA1 patient's data: (A) Cross-Validated Estimate of EPOCE, (B) difference in EPOCE and 95% tracking interval (top panel); (C) AUC Estimate and 95% pointwise Confidence Intervals (dashed lines), (D) Difference in AUC (solid line), 95% pointwise confidence intervals (dashed lines), and 95% Simultaneous Confidence Bands (dotted lines). CVPOL was the Cross-Validated Estimate of Expected Prognostic Observed Cross-Entropy (EPOCE).

Figure S1: Individual profiles SARA evolution over time according to the subtype disease.

Figure S2: Example of a virtual individual (SCA1 patients) predicted probability of survival within 5 years updated every 1 year from the JLC model (1-class 3-class) for longitudinal SARA score for a subject who died 4.9 years after the enrollment and had a disease duration of 5 years, CAG repeat number of 62, baseline SARA score = 22, baseline INAS score = 8, and repeated SARA score measurement denoted by x.

Table 1: Population Characteristics at Baseline

Characteristics	SCA1 (n=106^a)			SCA2 (n=146)			SCA3 (n=122)			SCA6 (n=87)		
	No.	%	Mean (SD)	No.	%	Mean (SD)	No.	%	Mean (SD)	No.	%	Mean (SD)
Male	66	62		68	47		61	50		48	55	
Age (years)			46 (12)			46 (14)			50 (12)			65 (11)
Age at onset (years)			37 (11)			35 (13)			38 (11)			55 (10)
Disease duration (years)			9 (5)			11 (6)			12 (6)			10 (6)
Repeat length expanded allele			49 (6)			40 (4)			71 (4)			22 (1)
Repeat length normal allele			30 (2)			22 (2)			22 (5)			12 (1)
SARA			14.8 (8.4)			15.7 (7.7)			14.1 (7.8)			15.0 (7.0)
Number of non-ataxia signs (INAS)			5 (2)			4 (2)			5 (2)			2 (2)
Death	32	30		33	23		26	21		11	13	

Abbreviations: SARA=scale for the assessment and rating of ataxia; SCA=spinocerebellar ataxia; Data are given as mean (SD) or number (%); ^a One patient (randomly selected) was removed to illustrate how joint model can be used to provide individual dynamic prediction.

Table 2: Multivariate Shared Random-Effect model (survival submodel) from SARA score According to the Genotype

Parameter	Estimate	SE	HR	95% CI		P
SCA1						
Age at baseline	0.032	0.015	1.03	1.00,	1.07	0.0383
Dysphagia (yes)	2.071	0.951	7.93	1.20,	52.27	0.0317
SARA (current value) ^b	0.244	0.039	1.28	1.18,	1.38	<.0001
SCA2						
Age at baseline	0.066	0.017	1.07	1.03,	1.11	0.0001
CAG (number repeats)	0.217	0.070	1.24	1.08,	1.43	0.0019
SARA (current value) ^b	0.156	0.029	1.17	1.10,	1.24	<0.0001
SCA3						
Age at baseline	0.213	0.249	1.24	0.76,	2.02	0.3937
CAG (number repeats)	-0.306	0.221	0.97	0.63,	1.49	0.8722
Interaction age and CAG	-0.003	0.004	0.99	0.98,	1.00	0.3915
SARA (current value) ^b	0.263	0.046	1.30	1.19,	1.42	<0.0001
SCA6						
Age at baseline	0.075	0.049	1.08	0.98,	1.19	0.1273
SARA (current value) ^b	0.229	0.064	1.26	1.11,	1.43	0.0004

Abbreviations: SARA, scale for the assessment and rating of ataxia; SCA, spinocerebellar ataxia; CAG, Cysteine Alanine Glycine;

^b the progression of SARA score was linear in all genotype, and was adjusted by the expanded repeat CAG for SCA1 patient; age at onset for SCA2; and no predictors for SCA3 and SCA6. The corresponding longitudinal submodel is given in appendix.

Table 3: Description of the Posterior Classification from SARA Joint Latent Class Model according to the Patient's Baseline Characteristics in SCA1 patients

Characteristics	Severe (n=13)			Intermediate (n=31)			Moderate (n=62)			P
	No.	%	Mean (SD)	No.	%	Mean (SD)	No.	%	Mean (SD)	
Male	7	53.8		20	66.7		39	60.9		0.7160
Age			52.1(13.7)			46.1(13.5)			44.8(10.6)	0.1324
Age at onset			35.3(11.5)			36.0(10.7)			37.7(10.3)	0.7241
Disease duration			16.8(4.5)			10.1(4.8)			7.1(4.1)	<.0001
Expanded CAG			48.3(7.3)			49.3(6.6)			46.8(4.6)	0.3241
Normal CAG			27.6(1.9)			29.7(2.1)			29.9(2.0)	0.0019
SARA			31.8(0.9)			18.9(0.7)			9.4(0.5)	<.0001
EQ-5D VAS			39.0(28.0)			51.8(18.5)			67.1(19.1)	0.0004
INAS			7.6(0.9)			6.0(0.4)			3.9(1.8)	<.0001
BMI			22.5(3.1)			23.6(3.9)			25.7(3.9)	0.0073
Cognitive impairment	7	53.8		7	23.3		6	9.4		0.0007
Urinary dysfunction	8	61.5		11	36.7		17	26.6		0.0601
Double vision	4	30.8		6	20.0		8	12.5		0.2371
Dysphagia	11	84.6		22	73.3		30	46.9		0.0253
Dystonia	4	30.8		8	22.8		1	1.6		<.0001
Death	10	83.3		17	51.5		4	6.6		<.0001

Abbreviations: SARA, scale for the assessment and rating of ataxia; SCA, spinocerebellar ataxia; CAG, Cysteine Alanine Glycine; INAS, Inventory of Non-Ataxia Signs, BMI, Body Mass Index, EQ-5D VAS, visual analogical scale (VAS) of the EuroQol five dimensions questionnaire (EQ-5D).

Figure 1:

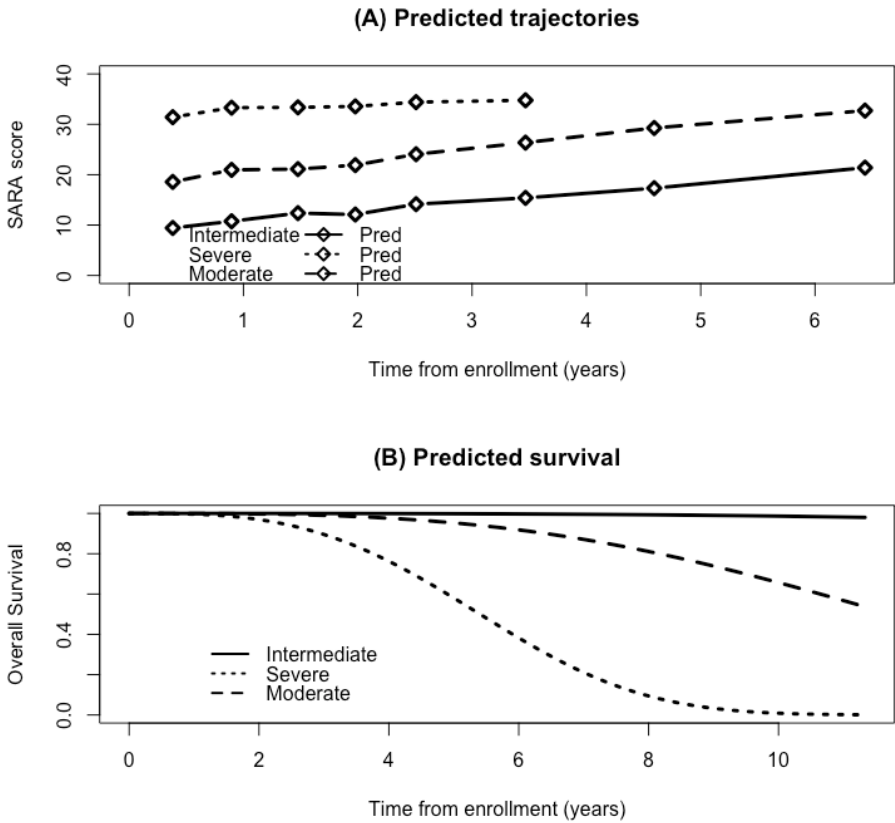
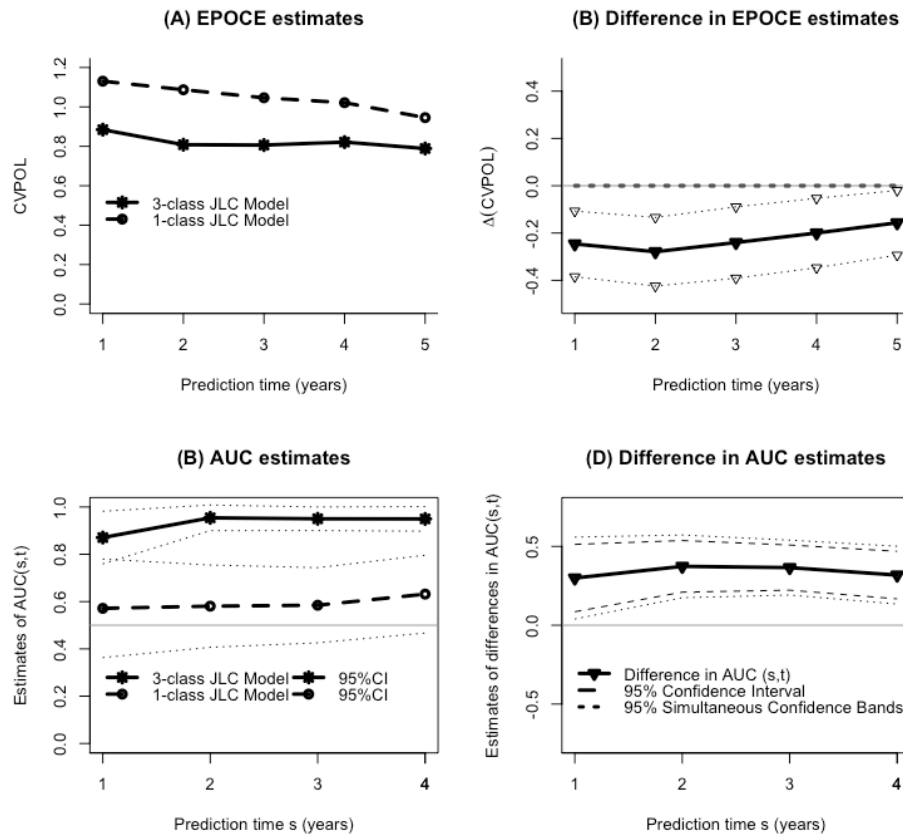


Figure 2:



CVPOL was the cross-validated estimate of expected prognostic observed cross-entropy (EPOCE).

Appendix

Figure S1:

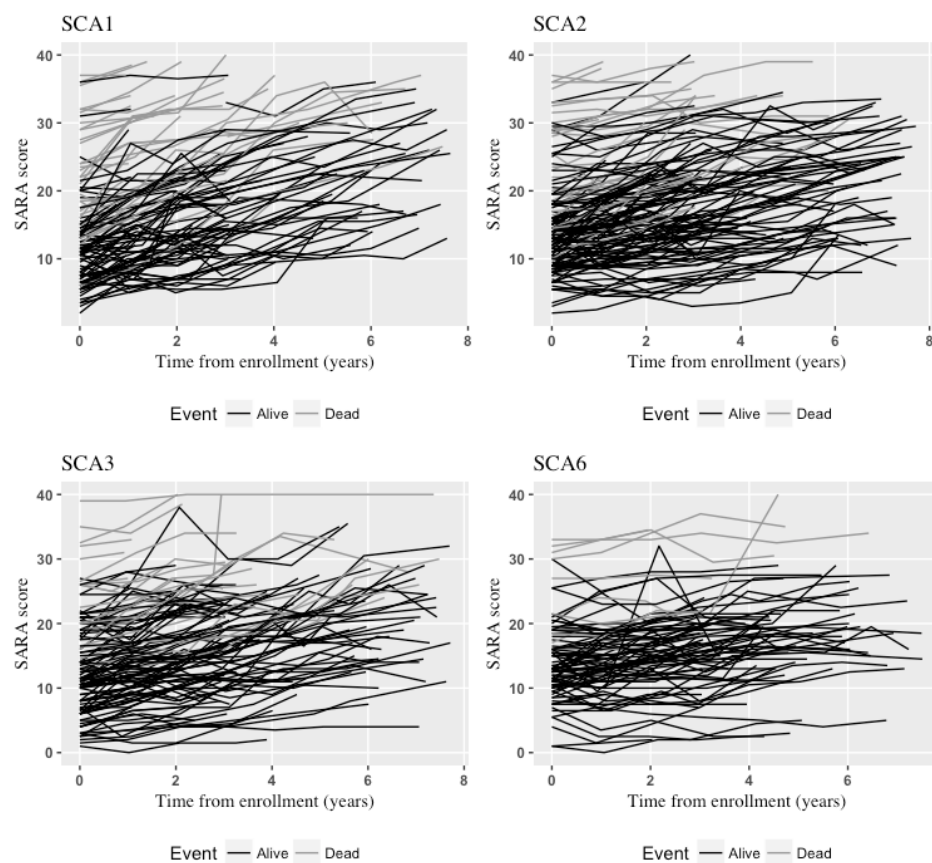
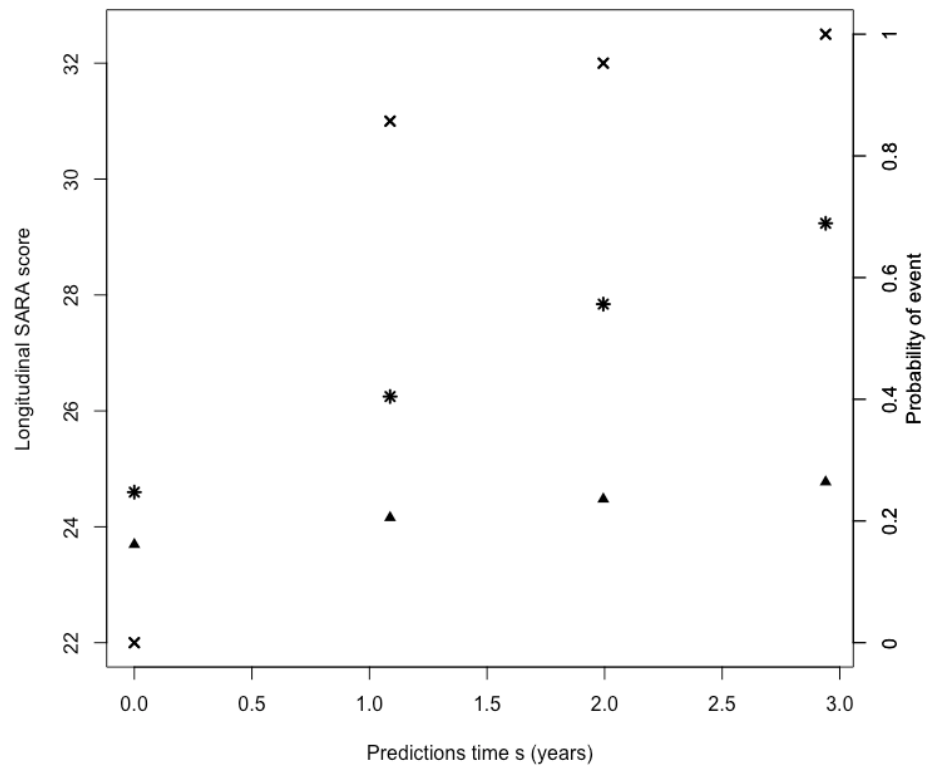


Figure S2:



A new patient was randomly selected and not included to build the joint model. Δ and $*$ represent the predicted probability of death within the time horizon of 5-year using all measures of SARA (denoted by x) collected till time at prediction s [0, 1.09, 1.99, and 2.94 years] from the 1-class JCL model and the 3-class JCL respectively.

Table S1: Choice of the Best Parameterization of the SARA Score Progression in SCA1, SCA2, SCA3, and SCA6

Model Parameterization	2*L	AIC	Level (SD)	Slope (SD)	2*L	AIC	Level (SD)	Slope (SD)
SCA1					SCA2			
Current value	2760.1	2786.1	0.24 (0.04)		3802.4	3838.4	0.17 (0.03)	
Slope	2807.9	2833.9		2.48 (0.55)	3834.8	3870.8		1.70 (0.59)
Baseline value	2801.6	2827.6	0.15 (0.02)		3809.3	3845.3	0.15 (0.03)	
Change from baseline value	2824.9	2850.9		0.14 (0.03)	3838.9	3874.9		0.20 (0.08)
Baseline and Change from baseline value	2759.9	2787.9	0.25 (0.04)	0.23 (0.04)	3802.4	3840.4	0.17 (0.03)	0.16 (0.07)
Current and slope	2759.3	2787.3	0.24 (0.04)	0.48 (0.54) ^c	3802.1	3840.1	0.17 (0.03)	0.30 (0.56) ^c
Cumulative Effects	2812.8	2838.8		0.01 (0.002)	3811.4	3847.4		0.02 (0.004)
Random-Effects	2811.5	2839.5	0.13 (0.03)	0.55 (0.51) ^c	3806.7	3844.7	0.17 (0.03)	0.93 (0.44) ^c
SCA3					SCA6			
Current value	3065.2	3097.2	0.28 (0.05)		2007.9	2031.9	0.20 (0.05)	
Slope	3109.5	3141.5		2.42 (0.66)	2027.6	2049.6		0.44 (0.49) ^c
Baseline value	3083.9	3115.9	0.24 (0.04)		2017.4	2039.4	0.15 (0.06)	
Change from baseline value	3114.3	3146.3		0.28 (0.08)	2027.5	2049.5		0.06 (0.06) ^c
Baseline and Change from baseline value	3065.1	3099.1	0.28 (0.05)	0.30 (0.08)	2008.5	2032.5	0.22 (0.06)	0.14 (0.07)
Current and slope	3064.6	3098.6	0.27 (0.05)	0.41 (0.56) ^c	2008.9	2032.9	0.22 (0.06)	0.57 (0.58) ^c
Cumulative Effects	3091.3	3123.3		0.03 (0.005)	2015.5	2037.5		0.02 (0.006)
Random-Effects	3073.1	3107.1	0.24 (0.04)	1.60 (0.57)	2027.6	2051.6		0.44 (0.49) ^c

^c $P > .05$; Log-likelihood (L), Akaike Information Criterion (AIC); for baseline and change from baseline value parameterization, the level (SD) and slope (SD) quantify the effect of baseline value and change from baseline value on survival respectively. For random-effects parameterization, the level (SD) and slope (SD) correspond to the effect of random intercept and random slope respectively. Based on the AIC, the best model parameterization was the one considering the current value.

Table S2: Multivariate Shared Random-Effect model (longitudinal submodel) from SARA score according to the genotype

Parameter	Estimate	SE	P
SCA1			
Intercept	8.6579	6.9508	0.2157
Slope	1.4675	0.6935	0.0367
CAG (number repeats)	0.1307	0.1451	0.3698
Interaction time and CAG	0.07435	0.01461	<.0001
SCA2			
Intercept	21.008	1.819	<0.0001
Slope	2.025	0.213	<0.0001
Age at onset	-0.152	0.048	0.0017
Interaction time and age at onset	-0.019	0.006	0.0006
SCA3			
Intercept	14.130	0.710	<0.0001
Slope	1.542	0.080	<0.0001
SCA6			
Intercept	14.394	0.711	<0.0001
Slope	0.865	0.097	<0.0001

Table S3: Summary of the Joint Model (JLCM and SREM) fits

Scores	Models	L	P	BIC	% class1	% class2	% class3	% class4	Entropy	ICL-BIC	ST (P)
SARA											
	JLCM										
	G = 1	-1413.2	13	2887.2	100.00				1	-	26.77 (<.0001)
	G = 2	-1386.3	19	2861.2	81.13	18.87			0.860	2881.7	13.89 (0.001)
	G = 3	-1369.5	25	2855.7	58.49	12.26	29.25		0.869	2886.2	1.068 (0.586)
	G = 4	-1362.8	31	2870.2	43.39	31.13	15.09	10.38	0.864	2910.0	1.472 (0.479)
	SREM										
	Y (t)	-1389.9	14	2840.5	-				-	-	-

Log-likelihood (L), number of parameters (P), Bayesian Information Criterion (BIC), number and proportion in each latent class (%), Entropy, Integrated Classification Likelihood with BIC approximation (ICL-BIC), Score statistic (ST) and associated p-value for the conditional independence test. For joint latent class model (JLCM), G=g indicates g latent classes, for shared random-effects model (SREM), Y (t) refer to the current level of the SARA score included as covariates in survival model. The optimal model with the lower ICL-BIC and satisfied conditional independence assumption include 3-class JLC model.

Table S4 : Posterior Classification

Mean posterior probability (AvePP)						
Scores	Group	N	%	Class 1	Class 2	Class 3
SARA						
	1	62	58.5	0.9445	0.0555	0.0000
	2	31	29.2	0.0785	0.8946	0.0269
	3	13	12.3	0.0000	0.0178	0.9822

Number (%) of subjects and mean posterior probabilities of belonging to each class according to the final posterior classification. For SARA progression, group=1 or Class=1 correspond to the moderate group in the main paper, group=2 or Class=2 correspond to the intermediate group, and group=3 or Class=3 correspond to the severe group.

7 Discussion générale

Au cours de cette thèse, nous avons abordé plusieurs défis soulevés par les recherches actuelles sur les ataxies spinocérébelleuses (SCA) notamment sur le plan clinique. L'essentiel de ce travail a porté sur les analyses en sous-groupes qui font l'hypothèse que la population d'étude est composée de plusieurs sous-populations d'évolution homogène d'un marqueur ou d'un score Y .

D'un point de vue clinique, ces travaux ont montré qu'en plus de l'hétérogénéité génétique, les ataxies spinocérébelleuses sont aussi cliniquement hétérogènes, grâce à l'identification de trois sous-groupes d'évolution du poids très corrélées à la progression de la maladie. De même, pour les SCA1, la modélisation conjointe de la survie et de la progression de l'ataxie a mis en évidence trois groupes homogènes de patients ayant les mêmes taux de progression et de risque de décès.

Compte tenu de la non plausibilité de l'hypothèse d'homogénéité de la population dans certaines situations (maladies génétiques ou du sujet âgé), les analyses en sous-groupes restent une alternative au modèle mixte standard. Elles présentent quelques avantages :

- mieux adapter le design des futures traitements, notamment la sélection et stratification des patients ;
- mieux identifier les groupes à risque (de progression et/ou de décès) permettant ainsi un meilleur suivi des patients.

Dans la suite, pour chacun des objectifs, après avoir passé en revue les principaux résultats, nous discuterons d'une part de leur impact sur la pratique clinique et d'autre part des principales limites.

7.1 Principaux résultats et leur impact

Premièrement, nous nous sommes intéressés à la relation entre la perte de poids et la progression de l'ataxie. Pour cela, nous avons étudié l'évolution du poids en fonction de l'âge à partir de la cohorte EUROSCA. Trois sous-groupes d'évolution du poids en fonction de l'âge ont été identifiés : évolution stable, croissante et décroissante. Nous avons montré que les patients qui baissent leur poids étaient plus sévères à l'inclusion (CAG long plus élevé) et que leur ataxie progressait plus rapidement par rapport aux autres groupes. Ce résultat suggère que, le poids qui est simple, accessible et facile à mesurer, pourrait être utilisé comme biomarqueur pour monitorer la progression de la maladie.

Pour identifier ces sous-groupes, nous avons utilisé les modèles mixtes à classes latentes proposés récemment [Commenges and Jacqmin-Gadda, 2016] comme une extension des modèles linéaires

mixtes. Ils ont l'avantage d'être plus souple en supposant l'hétérogénéité de la population d'étude, mais posent souvent des problèmes de convergence. Pour le choix des sous-groupes, nous avons utilisé un compromis entre la pertinence clinique et les critères statistiques (BIC plus petit). La probabilité moyenne *a posteriori* d'appartenir à une classe estimée pour chaque patient, variait entre 0.74 à 0.83 suggérant une bonne séparation des classes.

Notre deuxième travail portait sur l'estimation des taux de survie spécifique à chaque génotype et le développement d'un modèle pronostique de survie. Nos travaux sur la survie montrent deux résultats majeurs : la différence de survie entre les SCA caractérisant ainsi les patients SCA1 comme ceux qui ont le plus mauvais pronostic, et le développement d'un modèle prédictif spécifique à chaque génotype à partir des facteurs de risque indépendants de décès. Dans ce modèle, la sévérité de l'ataxie mesurée par le score SARA à l'inclusion était le facteur de risque le plus important de décès, soulignant la pertinence clinique et le pouvoir prédictif du SARA. Nous avons construit ce modèle prédictif en utilisant une méthodologie rigoureuse incluant le choix approprié de l'échelle de temps, et la validation interne du modèle par la procédure bootstrap. Ensuite, nous avons dérivé de ce modèle prédictif, un nomogramme présentant de très bonnes capacités discriminatives entre les patients vivants et décédés, et une excellente calibration. En plus, ce nomogramme inclut des prédicteurs cliniques de décès, simples et accessibles en pratique courante, et est d'une utilité très facile.

Enfin, nous avons utilisé la modélisation conjointe pour quantifier l'impact de la progression de l'ataxie à long-terme sur la survie (modèle conjoint à effets aléatoires partagés), et identifier des profils de patients partageant la même progression de l'ataxie et le même risque de décès (modèle conjoint à classes latentes). Nous avons montré que, la valeur courante du score SARA était la meilleure fonction de lien avec la survie. Ce résultat a été conforté par une analyse de sensibilité, où nous avons testé plusieurs structures de dépendances entre la progression de l'ataxie et la survie.

Uniquement chez les patients souffrant d'ataxie de type 1, trois profils de patients avec une progression de la maladie et un risque de décès identique ont été identifiés. Parmi eux, se distingue un petit groupe de 13 patients de très mauvais pronostic à l'inclusion, caractérisés par une rapide progression de l'ataxie associée à un risque de décès très élevé. Ensuite, nous avons prédit le décès à partir de la dynamique du score SARA en utilisant le modèle conjoint à classes latentes. Les prédictions dynamiques du modèle conjoint à 3-classes présentaient de très bonnes capacités discriminatives par rapport au modèle à 1-classe. Ces prédictions peuvent être très utiles pour le suivi des patients ou pour les décisions cliniques qui seront prise au cours du suivi.

Comparé au modèle conjoint à effets aléatoires partagés, le modèle conjoint à classes latentes présente, une meilleure adéquation aux données. La différence maximale de la log-vraisemblance entre les deux modèles conjoints était de 20.4. Comme cela a déjà été mentionné par d'autres auteurs [Proust-Lima et al., 2014], cette différence souligne la flexibilité du modèle conjoint à classes latentes à modéliser la progression de l'ataxie, la survie et leur dépendance. Cependant, cette flexibilité est associée à une augmentation du nombre de paramètres : 24 versus 14 respectivement pour les modèles conjoints à classes latentes et à effets aléatoires partagés, suggérant une certaine surestimation des paramètres.

A partir du nomogramme développé dans la troisième partie de la thèse, un score pronostique a été construit puis regroupé en 3 catégories : bon, intermédiaire et mauvais pronostic. Chez

les patients SCA1, nous avons comparé ce groupe à celui identifié par le modèle conjoint à 3-classes latentes. La concordance entre les deux méthodes était bonne (coefficient kappa = 0.74 [intervalle de confiance à 95%, 0.64 – 0.85]). Cependant, bien que les deux approches affichent un taux de concordance acceptable, le modèle conjoint présente plusieurs avantages. Tout d’abord, contrairement au modèle de Cox, le modèle conjoint permet de prendre en compte la mesure répétée du score SARA en corrigeant le biais induit par l’erreur de mesure aléatoire et la survenue des décès [Prentice, 1982]. Mais aussi, le modèle conjoint fournit une meilleure compréhension du lien entre le score clinique SARA répété et le risque de décès [Commenges and Jacqmin-Gadda, 2016].

7.2 Limites de la thèse

Malgré un recrutement prospectif de la cohorte EUROSCA, un nombre important de patients inclus par génotype et un suivi à plus de 10 ans, certaines limites demeurent dans ce travail.

La première limite concerne l’absence de données biologiques ou d’imagerie qui pourraient sans doute contribuer à l’enrichissement de nos différents modèles. Dans l’étude du lien entre le poids et l’ataxie, des mesures anthropologiques telles que l’épaisseur du pli cutané et le tour de taille qui pouvaient éventuellement servir de biomarqueur pour monitorer la progression de l’ataxie, ainsi que d’autres informations sur l’apport calorique ou le profil des lipides pouvant influencer le poids, n’étaient pas collectées dans EUROSCA. De même, les nomogrammes que nous avons développés pourraient être améliorés par l’ajout de biomarqueurs et de données d’imagerie absentes dans la cohorte EUROSCA.

La seconde limite est liée à l’absence de cohorte de validation qui nous auraient permis une validation externe des nomogrammes et ainsi que des outils de prédictions dynamiques afin qu’ils soient utilisables en pratique courante.

La troisième limite de notre étude concerne la possibilité d’un biais de survie résiduel. Nous avons certes mis à jour le statut vital pour minimiser ce biais en interrogeant les mairies de naissance ou les familles, mais nous ne sommes jamais sûrs de ne pas passer à coté de certains décès. Ce biais de survie pourrait avoir comme conséquence la surestimation des taux de survie.

La dernière limite est en rapport avec les sorties d’étude consécutives à l’extension de la cohorte au delà de 3 ans. Ces sorties ont un caractère informatif, car les patients avec un suivi court (≤ 3 ans) progressent plus rapidement que ceux avec suivi plus long (> 3 ans) [Jacobi et al., 2015]. Pour s’assurer que cette différence de suivi n’impactait pas nos estimations sur l’évolution du poids, nous avons réalisé une analyse de sensibilité en utilisant selon les recommandations actuelles, le modèle conjoint pour données longitudinales et de survie [Rizopoulos, 2012, Commenges and Jacqmin-Gadda, 2016]. Cette analyse de sensibilité suggérait une absence de biais dans l’estimation de nos paramètres [Diallo et al., 2017].

7.3 Conclusion et Perspective

Pour conclure, les résultats de nos travaux ont permis d’étendre les connaissances sur les caractéristiques biologiques des patients souffrants d’ataxie de types 1, 2, 3 et 6. Ils ont identifié

le génotype de type 1 (SCA1) comme étant le plus mauvais pronostique (progression plus rapide et survie plus courte) parmi les SCA. Ces résultats ont des implications directes dans le design des futurs essais cliniques, par exemple les nomogrammes pronostiques de survie pourraient être facilement utilisés pour la sélection et la stratification des patients.

Bien qu'il n'existe à ce jour aucun traitement efficace qui ralentirait la progression de l'ataxie, nos travaux ont permis une avancée majeure dans ce sens, notamment dans la compréhension de sa physiopathologie. Nos résultats ont contribué à l'initiation du projet READISCA (Clinical Trial Readiness for SCA1 and SCA3), sur lequel nous allons prochainement travailler, dont le but est la mise en place d'un essai thérapeutique. Le choix sur les SCA1 repose sur le fait qu'ils soient de plus mauvais pronostique et celui des SCA3 sur le fait qu'ils soient les plus prévalents. Ce projet est le fruit d'une collaboration US et Europe dont les objectifs seront de proposer un design adapté aux SCA, identifier un critère de jugement pertinent et adéquat pour évaluer l'efficacité des futurs traitements, et valider des biomarqueurs.

Bibliographie

- [Adanyeguh et al., 2015] Adanyeguh, I. M., Rinaldi, D., Henry, P.-G., Caillet, S., Valabregue, R., Durr, A., and Mochel, F. (2015). Triheptanoin improves brain energy metabolism in patients with Huntington disease. *Neurology*, 84(5) :490–495.
- [Almaguer-Mederos et al., 2013] Almaguer-Mederos, L. E., Aguilera Rodríguez, R., González Zaldivar, Y., Almaguer Gotay, D., Cuello Almarales, D., Laffita Mesa, J., Vázquez Mojena, Y., Zayas Feria, P., Auburger, G., Gispert, S., and Velásquez Pérez, L. (2013). Estimation of survival in spinocerebellar ataxia type 2 Cuban patients. *Clinical Genetics*, 83(3) :293–294.
- [Antenora et al., 2018] Antenora, A., Bruzzese, D., Lieto, M., Roca, A., Florio, M. T., Peluso, S., Saccà, F., De Michele, G., Santorelli, F. M., and Filla, A. (2018). Predictors of survival in spinocerebellar ataxia type 2 population from Southern Italy. *Neurological Sciences*.
- [Ashizawa et al., 2013] Ashizawa, T., Figueroa, K. P., Perlman, S. L., Gomez, C. M., Wilmot, G. R., Schmahmann, J. D., Ying, S. H., Zesiewicz, T. A., Paulson, H. L., Shakkottai, V. G., Bushara, K. O., Kuo, S.-H., Geschwind, M. D., Xia, G., Mazzoni, P., Krischer, J. P., Cuthbertson, D., Holbert, A., Ferguson, J. H., Pulst, S. M., and Subramony, S. (2013). Clinical characteristics of patients with spinocerebellar ataxias 1, 2, 3 and 6 in the US; a prospective observational study. *Orphanet Journal of Rare Diseases*, 8(1) :177.
- [Aziz et al., 2010] Aziz, N. A., Pijl, H., Frolich, M., Snel, M., Streefland, T. C. M., Roelfsema, F., and Roos, R. A. C. (2010). Systemic energy homeostasis in Huntington’s disease patients. *Journal of Neurology, Neurosurgery & Psychiatry*, 81(11) :1233–1237.
- [Aziz et al., 2008] Aziz, N. A., van der Burg, J., Landwehrmeyer, G. B., Brundin, P., Stijnen, T., EHDI Study Group, and Roos, R. A. (2008). Weight loss in Huntington disease increases with higher CAG repeat number. *Neurology*, 71(19) :1506–1513.
- [Balachandran et al., 2015] Balachandran, V. P., Gonen, M., Smith, J. J., and DeMatteo, R. P. (2015). Nomograms in oncology : more than meets the eye. *The Lancet Oncology*, 16(4) :e173–e180.
- [Basri et al., 2007] Basri, R., Yabe, I., Soma, H., and Sasaki, H. (2007). Spectrum and prevalence of autosomal dominant spinocerebellar ataxia in Hokkaido, the northern island of Japan : a study of 113 Japanese families. *Journal of Human Genetics*, 52(10) :848–855.
- [Bernhardson, 1975] Bernhardson, C. S. (1975). 375 : Type I Error Rates When Multiple Comparison Procedures Follow a Significant F Test of ANOVA. *Biometrics*, 31(1) :229.
- [Beunckens et al., 2008] Beunckens, C., Molenberghs, G., Verbeke, G., and Mallinckrodt, C. (2008). A Latent-Class Mixture Model for Incomplete Longitudinal Gaussian Data. *Biometrics*, 64(1) :96–105.
- [Blanche et al., 2015] Blanche, P., Proust-Lima, C., Loubère, L., Berr, C., Dartigues, J.-F., and Jacqmin-Gadda, H. (2015). Quantifying and comparing dynamic predictive accuracy of joint

- models for longitudinal marker and time-to-event in presence of censoring and competing risks : Comparing Dynamic Predictive Accuracy of Joint Models. *Biometrics*, 71(1) :102–113.
- [Boy et al., 2009] Boy, J., Schmidt, T., Wolburg, H., Mack, A., Nuber, S., Böttcher, M., Schmitt, I., Holzmann, C., Zimmermann, F., Servadio, A., and Riess, O. (2009). Reversibility of symptoms in a conditional mouse model of spinocerebellar ataxia type 3. *Human Molecular Genetics*, 18(22) :4282–4295.
- [Brusco et al., 2004] Brusco, A., Gellera, C., Cagnoli, C., Saluto, A., Castucci, A., Michielotto, C., Fetoni, V., Mariotti, C., Migone, N., Di Donato, S., and Taroni, F. (2004). Molecular Genetics of Hereditary Spinocerebellar Ataxia : Mutation Analysis of Spinocerebellar Ataxia Genes and CAG/CTG Repeat Expansion Detection in 225 Italian Families. *Archives of Neurology*, 61(5) :727.
- [Chalise et al., 2013] Chalise, P., Chicken, E., and McGee, D. (2013). Performance and Prediction for Varying Survival Time Scales. *Communications in Statistics - Simulation and Computation*, 42(3) :636–649.
- [Chalise et al., 2016] Chalise, P., Chicken, E., and McGee, D. (2016). Time Scales in Epidemiological Analysis : An Empirical Comparison. *International Journal of Statistics and Probability*, 5(3) :91.
- [Commenges and Jacqmin-Gadda, 2016] Commenges, D. and Jacqmin-Gadda, H. (2016). *Dynamical biostatistical models*. Number 86 in Chapman & Hall/CRC biostatistics series. CRC Press/Taylor & Francis, Boca Raton.
- [Commenges et al., 2012] Commenges, D., Liqueur, B., and Proust-Lima, C. (2012). Choice of Prognostic Estimators in Joint Models by Estimating Differences of Expected Conditional Kullback-Leibler Risks. *Biometrics*, 68(2) :380–387.
- [Cox, 1957] Cox, D. R. (1957). Note on Grouping. *Journal of the American Statistical Association*, 52(280) :543.
- [Dantan et al., 2008] Dantan, E., Proust-Lima, C., Letenneur, L., and Jacqmin-Gadda, H. (2008). Pattern mixture models and latent class models for the analysis of multivariate longitudinal data with informative dropouts. *The International Journal of Biostatistics*, 4(1) :Article 14.
- [Diallo et al., 2018] Diallo, A., Jacobi, H., Cook, A., Labrum, R., Durr, A., Brice, A., Charles, P., Marelli, C., Mariotti, C., Nanetti, L., Panzeri, M., Rakowicz, M., Sobanska, A., Sulek, A., Schmitz-Hübsch, T., Schöls, L., Hengel, H., Melegh, B., Filla, A., Antenora, A., Infante, J., Berciano, J., van de Warrenburg, B. P., Timmann, D., Boesch, S., Pandolfo, M., Schulz, J. B., Bauer, P., Giunti, P., Kang, J.-S., Klockgether, T., and Tezenas du Montcel, S. (2018). Survival in patients with spinocerebellar ataxia types 1, 2, 3, and 6 (EUROSCA) : a longitudinal cohort study. *The Lancet Neurology*, 17(4) :327–334.
- [Diallo et al., 2017] Diallo, A., Jacobi, H., Schmitz-Hübsch, T., Cook, A., Labrum, R., Durr, A., Brice, A., Charles, P., Marelli, C., Mariotti, C., Nanetti, L., Panzeri, M., Rakowicz, M., Sobanska, A., Sulek, A., Schöls, L., Hengel, H., Melegh, B., Filla, A., Antenora, A., Infante, J., Berciano, J., van de Warrenburg, B. P., Timmann, D., Boesch, S., Pandolfo, M., Schulz, J. B., Bauer, P., Giunti, P., Baliko, L., Parkinson, M. H., Kang, J.-S., Klockgether, T., and Tezenas du Montcel, S. (2017). Body Mass Index Decline Is Related to Spinocerebellar Ataxia Disease Progression. *Movement Disorders Clinical Practice*, 4(5) :689–697.
- [Duarte-Silva et al., 2014] Duarte-Silva, S., Neves-Carvalho, A., Soares-Cunha, C., Teixeira-Castro, A., Oliveira, P., Silva-Fernandes, A., and Maciel, P. (2014). Lithium Chloride Therapy

- Fails to Improve Motor Function in a Transgenic Mouse Model of Machado-Joseph Disease. *The Cerebellum*, 13(6) :713–727.
- [Durr, 2010] Durr, A. (2010). Autosomal dominant cerebellar ataxias : polyglutamine expansions and beyond. *The Lancet Neurology*, 9(9) :885–894.
- [Durr et al., 1996] Durr, A., Stevanin, G., Cancel, G., Duyckaerts, C., Abbas, N., Didierjean, O., Chneiweiss, H., Benomar, A., Lyon-Caen, O., Julien, J., Serdaru, M., Penet, C., Agid, Y., and Brice, A. (1996). Spinocerebellar ataxia 3 and machado-joseph disease : Clinical, molecular, and neuropathological features. *Annals of Neurology*, 39(4) :490–499.
- [Erichsen et al., 2009] Erichsen, A. K., Koht, J., Stray-Pedersen, A., Abdelnoor, M., and Tallaksen, C. M. E. (2009). Prevalence of hereditary ataxia and spastic paraplegia in southeast Norway : a population-based study. *Brain*, 132(6) :1577–1588.
- [EuroQol Group, 1990] EuroQol Group (1990). EuroQol—a new facility for the measurement of health-related quality of life. *Health Policy (Amsterdam, Netherlands)*, 16(3) :199–208.
- [Faruq et al., 2009] Faruq, M., Scaria, V., Singh, I., Tyagi, S., Srivastava, A. K., and Mukerji, M. (2009). SCA-LSVD : A repeat-oriented locus-specific variation database for genotype to phenotype correlations in spinocerebellar ataxias. *Human Mutation*, 30(7) :1037–1042.
- [Fogel and Perlman, 2006] Fogel, B. L. and Perlman, S. (2006). An approach to the patient with late-onset cerebellar ataxia. *Nature Clinical Practice Neurology*, 2(11) :629–635.
- [Globas et al., 2008] Globas, C., du Montcel, S. T., Baliko, L., Boesch, S., Depondt, C., DiDonato, S., Durr, A., Filla, A., Klockgether, T., Mariotti, C., Melegh, B., Rakowicz, M., Ribai, P., Rola, R., Schmitz-Hubsch, T., Szymanski, S., Timmann, D., Van de Warrenburg, B. P., Bauer, P., and Schols, L. (2008). Early symptoms in spinocerebellar ataxia type 1, 2, 3, and 6. *Movement Disorders*, 23(15) :2232–2238.
- [Group, 1996] Group, H. S. (1996). Unified Huntington’s disease rating scale : Reliability and consistency. *Movement Disorders*, 11(2) :136–142.
- [Group, 1993] Group, T. H. D. C. R. (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington’s disease chromosomes. The Huntington’s Disease Collaborative Research Group. *Cell*, 72(6) :971–983.
- [Guerrini et al., 2004] Guerrini, L., Lolli, F., Ginestroni, A., Belli, G., Della Nave, R., Tessa, C., Foresti, S., Cosottini, M., Piacentini, S., Salvi, F., Plasmatti, R., De Grandis, D., Siciliano, G., Filla, A., and Mascalchi, M. (2004). Brainstem neurodegeneration correlates with clinical dysfunction in SCA1 but not in SCA2. A quantitative volumetric, diffusion and proton spectroscopy MR study. *Brain : A Journal of Neurology*, 127(Pt 8) :1785–1795.
- [Han et al., 2007a] Han, J., Slate, E. H., and Peña, E. A. (2007a). Parametric latent class joint model for a longitudinal biomarker and recurrent events. *Statistics in Medicine*, 26(29) :5285–5302.
- [Han et al., 2007b] Han, J., Slate, E. H., and Peña, E. A. (2007b). Parametric latent class joint model for a longitudinal biomarker and recurrent events. *Statistics in Medicine*, 26(29) :5285–5302.
- [Harding, 1983] Harding, A. E. (1983). Classification of the hereditary ataxias and paraplegias. *Lancet (London, England)*, 1(8334) :1151–1155.
- [Hayashi et al., 2012] Hayashi, Y., Xiao, L., Suzuki, A., Blum, M. A., Sabloff, B., Taketa, T., Maru, D. M., Welsh, J., Lin, S. H., Weston, B., Lee, J. H., Bhutani, M. S., Hofstetter, W. L., Swisher,

- S. G., and Ajani, J. A. (2012). A nomogram associated with high probability of malignant nodes in the surgical specimen after trimodality therapy of patients with oesophageal cancer. *European Journal of Cancer*, 48(18) :3396–3404.
- [Hedeker and Gibbons, 1997] Hedeker, D. and Gibbons, R. D. (1997). Application of random-effects pattern-mixture models for missing data in longitudinal studies. *Psychological Methods*, 2(1) :64–78.
- [Hübener et al., 2012] Hübener, J., Casadei, N., Teismann, P., Seeliger, M. W., Björkqvist, M., von Hörsten, S., Riess, O., and Nguyen, H. P. (2012). Automated Behavioral Phenotyping Reveals Presymptomatic Alterations in a SCA3 Genetrap Mouse Model. *Journal of Genetics and Genomics*, 39(6) :287–299.
- [Iasonos et al., 2008] Iasonos, A., Schrag, D., Raj, G. V., and Panageas, K. S. (2008). How To Build and Interpret a Nomogram for Cancer Prognosis. *Journal of Clinical Oncology*, 26(8) :1364–1370.
- [Iizuka et al., 2015a] Iizuka, A., Nakamura, K., and Hirai, H. (2015a). Long-term oral administration of the NMDA receptor antagonist memantine extends life span in spinocerebellar ataxia type 1 knock-in mice. *Neuroscience Letters*, 592 :37–41.
- [Iizuka et al., 2015b] Iizuka, A., Nakamura, K., and Hirai, H. (2015b). Long-term oral administration of the NMDA receptor antagonist memantine extends life span in spinocerebellar ataxia type 1 knock-in mice. *Neuroscience Letters*, 592 :37–41.
- [Jacobi et al., 2011] Jacobi, H., Bauer, P., Giunti, P., Labrum, R., Sweeney, M. G., Charles, P., Durr, A., Marelli, C., Globas, C., Linnemann, C., Schols, L., Rakowicz, M., Rola, R., Zdzienicka, E., Schmitz-Hubsch, T., Fancellu, R., Mariotti, C., Tomasello, C., Baliko, L., Melegh, B., Filla, A., Rinaldi, C., van de Warrenburg, B. P., Verstappen, C. C. P., Szymanski, S., Berciano, J., Infante, J., Timmann, D., Boesch, S., Hering, S., Depondt, C., Pandolfo, M., Kang, J.-S., Ratzka, S., Schulz, J., Tezenas du Montcel, S., and Klockgether, T. (2011). The natural history of spinocerebellar ataxia type 1, 2, 3, and 6 : A 2-year follow-up study. *Neurology*, 77(11) :1035–1041.
- [Jacobi et al., 2015] Jacobi, H., du Montcel, S. T., Bauer, P., Giunti, P., Cook, A., Labrum, R., Parkinson, M. H., Durr, A., Brice, A., Charles, P., Marelli, C., Mariotti, C., Nanetti, L., Panzeri, M., Rakowicz, M., Sulek, A., Sobanska, A., Schmitz-Hübsch, T., Schöls, L., Hengel, H., Baliko, L., Melegh, B., Filla, A., Antenora, A., Infante, J., Berciano, J., van de Warrenburg, B. P., Timmann, D., Szymanski, S., Boesch, S., Kang, J.-S., Pandolfo, M., Schulz, J. B., Molho, S., Diallo, A., and Klockgether, T. (2015). Long-term disease progression in spinocerebellar ataxia types 1, 2, 3, and 6 : a longitudinal cohort study. *The Lancet Neurology*, 14(11) :1101–1108.
- [Jacobi et al., 2013a] Jacobi, H., Rakowicz, M., Rola, R., Fancellu, R., Mariotti, C., Charles, P., Dürr, A., Küper, M., Timmann, D., Linnemann, C., Schöls, L., Kaut, O., Schaub, C., Filla, A., Baliko, L., Melegh, B., Kang, J.-S., Giunti, P., van de Warrenburg, B. P. C., Fimmers, R., and Klockgether, T. (2013a). Inventory of Non-Ataxia Signs (INAS) : Validation of a New Clinical Assessment Instrument. *The Cerebellum*, 12(3) :418–428.
- [Jacobi et al., 2013b] Jacobi, H., Reetz, K., du Montcel, S. T., Bauer, P., Mariotti, C., Nanetti, L., Rakowicz, M., Sulek, A., Durr, A., Charles, P., Filla, A., Antenora, A., Schöls, L., Schicks, J., Infante, J., Kang, J.-S., Timmann, D., Fabio, R. D., Masciullo, M., Baliko, L., Melegh, B., Boesch, S., Bürk, K., Peltz, A., Schulz, J. B., Dufaure-Garé, I., and Klockgether, T. (2013b). Biological and clinical characteristics of individuals at risk for spinocerebellar ataxia types 1, 2, 3, and 6 in the longitudinal RISCA study : analysis of baseline data. *The Lancet Neurology*, 12(7) :650–658.

- [Jacqmin-Gadda et al., 2010] Jacqmin-Gadda, H., Proust-Lima, C., Taylor, J. M., and Comenges, D. (2010). Score Test for Conditional Independence Between Longitudinal Outcome and Time to Event Given the Classes in the Joint Latent Class Model. *Biometrics*, 66(1) :11–19.
- [Jafar-Nejad et al., 2011] Jafar-Nejad, P., Ward, C. S., Richman, R., Orr, H. T., and Zoghbi, H. Y. (2011). Regional rescue of spinocerebellar ataxia type 1 phenotypes by *14-3-3* haploinsufficiency in mice underscores complex pathogenicity in neurodegeneration. *Proceedings of the National Academy of Sciences*, 108(5) :2142–2147.
- [Julio et al., 2010] Julio, M.-B., Gilberto, S.-C., Andres, M. G., Mario, E. B., and Luis, V.-P. (2010). Heart rate variability in type 2 spinocerebellar ataxia. *Acta Neurologica Scandinavica*.
- [Keller and Bastian, 2014a] Keller, J. L. and Bastian, A. J. (2014a). A Home Balance Exercise Program Improves Walking in People With Cerebellar Ataxia. *Neurorehabilitation and Neural Repair*, 28(8) :770–778.
- [Keller and Bastian, 2014b] Keller, J. L. and Bastian, A. J. (2014b). A Home Balance Exercise Program Improves Walking in People With Cerebellar Ataxia. *Neurorehabilitation and Neural Repair*, 28(8) :770–778.
- [Kieling et al., 2007] Kieling, C., Prestes, P., Saraiva-Pereira, M., and Jardim, L. (2007). Survival estimates for patients with Machado-Joseph disease (SCA3). *Clinical Genetics*, 72(6) :543–545.
- [Kim et al., 2014] Kim, D.-Y., Shim, S.-H., Kim, S.-O., Lee, S.-W., Park, J.-Y., Suh, D.-S., Kim, J.-H., Kim, Y.-M., Kim, Y.-T., and Nam, J.-H. (2014). Preoperative nomogram for the identification of lymph node metastasis in early cervical cancer. *British Journal of Cancer*, 110(1) :34–41.
- [Klockgether, 2008] Klockgether, T. (2008). The clinical diagnosis of autosomal dominant spinocerebellar ataxias. *The Cerebellum*, 7(2) :101–105.
- [Klockgether, 2010] Klockgether, T. (2010). Sporadic ataxia with adult onset : classification and diagnostic criteria. *The Lancet Neurology*, 9(1) :94–104.
- [Klockgether et al., 1998] Klockgether, T., Lüdtke, R., Kramer, B., Abele, M., Bürk, K., Schöls, L., Riess, O., Laccone, F., Boesch, S., Lopes-Cendes, I., Brice, A., Inzelberg, R., Zilber, N., and Dichgans, J. (1998). The natural history of degenerative ataxia : a retrospective study in 466 patients. *Brain : A Journal of Neurology*, 121 (Pt 4) :589–600.
- [Kordasiewicz and Gomez, 2007] Kordasiewicz, H. B. and Gomez, C. M. (2007). Molecular pathogenesis of spinocerebellar ataxia type 6. *Neurotherapeutics*, 4(2) :285–294.
- [Kraft et al., 2005] Kraft, S., Furtado, S., Ranaway, R., Parboosingh, J., Bleoo, S., McElligott, K., Bridge, P., Spacey, S., Das, S., and Suchowersky, O. (2005). Adult onset spinocerebellar ataxia in a Canadian movement disorders clinic. *The Canadian Journal of Neurological Sciences. Le Journal Canadien Des Sciences Neurologiques*, 32(4) :450–458.
- [Laird and Ware, 1982] Laird, N. M. and Ware, J. H. (1982). Random-Effects Models for Longitudinal Data. *Biometrics*, 38(4) :963.
- [Lamberink et al., 2017] Lamberink, H. J., Otte, W. M., Geerts, A. T., Pavlovic, M., Ramos-Lizana, J., Marson, A. G., Overweg, J., Sauma, L., Specchio, L. M., Tennison, M., Cardoso, T. M. O., Shinnar, S., Schmidt, D., Geleijns, K., and Braun, K. P. J. (2017). Individualised prediction model of seizure recurrence and long-term outcomes after withdrawal of antiepileptic drugs in seizure-free patients : a systematic review and individual participant data meta-analysis. *The Lancet Neurology*, 16(7) :523–531.

- [Lee et al., 2011] Lee, Y.-c., Liao, Y.-c., Wang, P.-s., Lee, I.-H., Lin, K.-p., and Soong, B.-w. (2011). Comparison of cerebellar ataxias : A three-year prospective longitudinal assessment. *Movement Disorders*, 26(11) :2081–2087.
- [Leslie et al., 2018] Leslie, W. D., Shevroja, E., Johansson, H., McCloskey, E. V., Harvey, N. C., Kanis, J. A., and Hans, D. (2018). Risk-equivalent T-score adjustment for using lumbar spine trabecular bone score (TBS) : the Manitoba BMD registry. *Osteoporosis International*, 29(3) :751–758.
- [Liang et al., 2016] Liang, L., Chen, T., and Wu, Y. (2016). The electrophysiology of spinocerebellar ataxias. *Neurophysiologie Clinique/Clinical Neurophysiology*, 46(1) :27–34.
- [Lin et al., 2002] Lin, H., Turnbull, B. W., McCulloch, C. E., and Slate, E. H. (2002). Latent Class Models for Joint Analysis of Longitudinal Biomarker and Event Process Data : Application to Longitudinal Prostate-Specific Antigen Readings and Prostate Cancer. *Journal of the American Statistical Association*, 97(457) :53–65.
- [Lin et al., 2013] Lin, N. X., Logan, S., and Henley, W. E. (2013). Bias and Sensitivity Analysis When Estimating Treatment Effects from the Cox Model with Omitted Covariates : Bias and Sensitivity Analysis for the Cox Model. *Biometrics*, 69(4) :850–860.
- [Lin et al., 2000] Lin, X., Antalffy, B., Kang, D., Orr, H. T., and Zoghbi, H. Y. (2000). Polyglutamine expansion down-regulates specific neuronal genes before pathologic changes in SCA1. *Nature Neuroscience*, 3(2) :157–163.
- [Little, 1995] Little, R. J. A. (1995). Modeling the Drop-Out Mechanism in Repeated-Measures Studies. *Journal of the American Statistical Association*, 90(431) :1112.
- [Liu et al., 2009] Liu, J., Tang, T.-S., Tu, H., Nelson, O., Herndon, E., Huynh, D. P., Pulst, S. M., and Bezprozvanny, I. (2009). Deranged Calcium Signaling and Neurodegeneration in Spinocerebellar Ataxia Type 2. *Journal of Neuroscience*, 29(29) :9148–9162.
- [López-Bastida et al., 2008] López-Bastida, J., Perestelo-Pérez, L., Montón-álvarez, F., and Serrano-Aguilar, P. (2008). Social economic costs and health-related quality of life in patients with degenerative cerebellar ataxia in Spain : Costs and Quality of Life in Ataxias. *Movement Disorders*, 23(2) :212–217.
- [Maas et al., 2015] Maas, R. P. P. W. M., van Gaalen, J., Klockgether, T., and van de Warrenburg, B. P. C. (2015). The preclinical stage of spinocerebellar ataxias. *Neurology*, 85(1) :96–103.
- [Maruyama et al., 2002] Maruyama, H., Izumi, Y., Morino, H., Oda, M., Toji, H., Nakamura, S., and Kawakami, H. (2002). Difference in disease-free survival curve and regional distribution according to subtype of spinocerebellar ataxia : A study of 1,286 Japanese patients. *American Journal of Medical Genetics*, 114(5) :578–583.
- [Matilla-Dueñas et al., 2010] Matilla-Dueñas, A., Sánchez, I., Corral-Juan, M., Dávalos, A., Alvarez, R., and Latorre, P. (2010). Cellular and Molecular Pathways Triggering Neurodegeneration in the Spinocerebellar Ataxias. *The Cerebellum*, 9(2) :148–166.
- [Mochel et al., 2007] Mochel, F., Charles, P., Seguin, F., Barritault, J., Coussieu, C., Perin, L., Le Bouc, Y., Gervais, C., Carcelain, G., Vassault, A., Feingold, J., Rabier, D., and Durr, A. (2007). Early Energy Deficit in Huntington Disease : Identification of a Plasma Biomarker Traceable during Disease Progression. *PLoS ONE*, 2(7) :e647.
- [Molenberghs et al., 2015] Molenberghs, G., Fitzmaurice, G. M., Kenward, M. G., Tsiatis, A. A., and Verbeke, G., editors (2015). *Handbook of missing data methodology*. Chapman & Hall/CRC handbooks of modern statistical methods. CRC Press, Taylor & Francis Group, Boca Raton.

- [Monin et al., 2015] Monin, M.-L., Tezenas du Montcel, S., Marelli, C., Cazeneuve, C., Charles, P., Tallaksen, C., Forlani, S., Stevanin, G., Brice, A., and Durr, A. (2015). Survival and severity in dominant cerebellar ataxias. *Annals of Clinical and Translational Neurology*, 2(2) :202–207.
- [Moseley et al., 1998] Moseley, M. L., Benzow, K. A., Schut, L. J., Bird, T. D., Gomez, C. M., Barkhaus, P. E., Blindauer, K. A., Labuda, M., Pandolfo, M., Koob, M. D., and Ranum, L. P. (1998). Incidence of dominant spinocerebellar and Friedreich triplet repeats among 361 ataxia families. *Neurology*, 51(6) :1666–1671.
- [Muthén and Shedden, 1999] Muthén, B. and Shedden, K. (1999). Finite mixture modeling with mixture outcomes using the EM algorithm. *Biometrics*, 55(2) :463–469.
- [Mähler et al., 2014] Mähler, A., Steiniger, J., Endres, M., Paul, F., Boschmann, M., and Doss, S. (2014). Increased Catabolic State in Spinocerebellar Ataxia Type 1 Patients. *The Cerebellum*, 13(4) :440–446.
- [Netravathi et al., 2009] Netravathi, M., Sathyaprabha, T. N., Jayalaxmi, K., Datta, P., Nirmala, M., and Pal, P. K. (2009). A comparative study of cardiac dysautonomia in autosomal dominant spinocerebellar ataxias and idiopathic sporadic ataxias. *Acta Neurologica Scandinavica*, 120(3) :204–209.
- [Orozco et al., 1989] Orozco, G., Estrada, R., Perry, T. L., Araña, J., Fernandez, R., Gonzalez-Quevedo, A., Galarraga, J., and Hansen, S. (1989). Dominantly inherited olivopontocerebellar atrophy from eastern Cuba. Clinical, neuropathological, and biochemical findings. *Journal of the Neurological Sciences*, 93(1) :37–50.
- [Paganoni et al., 2011] Paganoni, S., Deng, J., Jaffa, M., Cudkowicz, M. E., and Wills, A.-M. (2011). Body mass index, not dyslipidemia, is an independent predictor of survival in amyotrophic lateral sclerosis. *Muscle & Nerve*, 44(1) :20–24.
- [Pedroso et al., 2017] Pedroso, J. L., Braga-Neto, P., Escorcio-Bezerra, M. L., Abrahão, A., de Albuquerque, M. V. C., Filho, F. M. R., de Souza, P. V. S., de Rezende Pinto, W. B. V., Borges, F. R. P., Saraiva-Pereira, M. L., Jardim, L. B., and Barsottini, O. G. P. (2017). Non-motor and Extracerebellar Features in Spinocerebellar Ataxia Type 2. *The Cerebellum*, 16(1) :34–39.
- [Pedroso et al., 2013] Pedroso, J. L., França, M. C., Braga-Neto, P., D’Abreu, A., Saraiva-Pereira, M. L., Saute, J. A., Teive, H. A., Caramelli, P., Jardim, L. B., Lopes-Cendes, I., and Barsottini, O. G. P. (2013). Nonmotor and extracerebellar features in Machado-Joseph disease : A review : Extracerebellar Features in Machado-Joseph Disease. *Movement Disorders*, 28(9) :1200–1208.
- [Prentice, 1982] Prentice, R. L. (1982). Covariate Measurement Errors and Parameter Estimation in a Failure Time Regression Model. *Biometrika*, 69(2) :331.
- [Proust-Lima et al., 2017] Proust-Lima, C., Philipps, V., and Lique, B. (2017). Estimation of Extended Mixed Models Using Latent Classes and Latent Processes : The R Package **lcmm**. *Journal of Statistical Software*, 78(2).
- [Proust-Lima et al., 2014] Proust-Lima, C., Séne, M., Taylor, J. M., and Jacqmin-Gadda, H. (2014). Joint latent class models for longitudinal and time-to-event data : A review. *Statistical Methods in Medical Research*, 23(1) :74–90.
- [Proust-Lima and Taylor, 2009] Proust-Lima, C. and Taylor, J. M. G. (2009). Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of posttreatment PSA : a joint modeling approach. *Biostatistics*, 10(3) :535–549.
- [Rizopoulos, 2012] Rizopoulos, D. (2012). *Joint models for longitudinal and time-to-event data : with applications in R*. Number 6 in Chapman & Hall/CRC biostatistics series. CRC Press, Boca Raton.

- [Robins et al., 1995] Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1995). Analysis of Semiparametric Regression Models for Repeated Outcomes in the Presence of Missing Data. *Journal of the American Statistical Association*, 90(429) :106.
- [Roy, 2003] Roy, J. (2003). Modeling longitudinal data with nonignorable dropouts using a latent dropout class model. *Biometrics*, 59(4) :829–836.
- [Royston and Altman, 2013] Royston, P. and Altman, D. G. (2013). External validation of a Cox prognostic model : principles and methods. *BMC Medical Research Methodology*, 13(1).
- [Rüb et al., 2006] Rüb, U., Brunt, E. R., Petrasch-Parwez, E., Schöls, L., Theegarten, D., Auburger, G., Seidel, K., Schultz, C., Gierga, K., Paulson, H., van Broeckhoven, C., Deller, T., and de Vos, R. A. I. (2006). Degeneration of ingestion-related brainstem nuclei in spinocerebellar ataxia type 2, 3, 6 and 7. *Neuropathology and Applied Neurobiology*, 32(6) :635–649.
- [Saute et al., 2011] Saute, J. A. M., da Silva, A. C. F., Muller, A. P., Hansel, G., de Mello, A. S., Maeda, F., Vedolin, L., Saraiva-Pereira, M. L., Souza, D. O., Arpa, J., Torres-Aleman, I., Portela, L. V. C., and Jardim, L. B. (2011). Serum insulin-like system alterations in patients with spinocerebellar ataxia type 3 : Insulin-Like System in SCA3. *Movement Disorders*, 26(4) :731–735.
- [Saute et al., 2012] Saute, J. A. M., da Silva, A. C. F., Souza, G. N., Russo, A. D., Donis, K. C., Vedolin, L., Saraiva-Pereira, M. L., Portela, L. V. C., and Jardim, L. B. (2012). Body Mass Index is Inversely Correlated with the Expanded CAG Repeat Length in SCA3/MJD Patients. *The Cerebellum*, 11(3) :771–774.
- [Schmitz-Hubsch et al., 2008a] Schmitz-Hubsch, T., Coudert, M., Bauer, P., Giunti, P., Globas, C., Baliko, L., Filla, A., Mariotti, C., Rakowicz, M., Charles, P., Ribai, P., Szymanski, S., Infante, J., van de Warrenburg, B., Durr, A., Timmann, D., Boesch, S., Fancellu, R., Rola, R., Depondt, C., Schols, L., Zdienicka, E., Kang, J. S., Dohlinger, S., Kremer, B., Stephenson, D. A., Melegh, B., Pandolfo, M., di Donato, S., du Montcel, S. T., and Klockgether, T. (2008a). Spinocerebellar ataxia types 1, 2, 3, and 6 : Disease severity and nonataxia symptoms. *Neurology*, 71(13) :982–989.
- [Schmitz-Hubsch et al., 2006] Schmitz-Hubsch, T., du Montcel, S. T., Baliko, L., Berciano, J., Boesch, S., Depondt, C., Giunti, P., Globas, C., Infante, J., Kang, J. S., Kremer, B., Mariotti, C., Melegh, B., Pandolfo, M., Rakowicz, M., Ribai, P., Rola, R., Schols, L., Szymanski, S., van de Warrenburg, B. P., Durr, A., and Klockgether, T. (2006). Scale for the assessment and rating of ataxia : Development of a new clinical scale. *Neurology*, 66(11) :1717–1720.
- [Schmitz-Hubsch et al., 2010] Schmitz-Hubsch, T., Fimmers, R., Rakowicz, M., Rola, R., Zdzienicka, E., Fancellu, R., Mariotti, C., Linnemann, C., Schols, L., Timmann, D., Filla, A., Salvatore, E., Infante, J., Giunti, P., Labrum, R., Kremer, B., van de Warrenburg, B., Baliko, L., Melegh, B., Depondt, C., Schulz, J., du Montcel, S. T., and Klockgether, T. (2010). Responsiveness of different rating instruments in spinocerebellar ataxia patients. *Neurology*, 74(8) :678–684.
- [Schmitz-Hubsch et al., 2008b] Schmitz-Hubsch, T., Giunti, P., Stephenson, D. A., Globas, C., Baliko, L., Sacca, F., Mariotti, C., Rakowicz, M., Szymanski, S., Infante, J., van de Warrenburg, B., Timmann, D., Fancellu, R., Rola, R., Depondt, C., Schols, L., Zdzienicka, E., Kang, J. S., Dohlinger, S., Kremer, B., Melegh, B., Filla, A., and Klockgether, T. (2008b). SCA Functional Index : A useful compound performance measure for spinocerebellar ataxia. *Neurology*, 71(7) :486–492.
- [Schulz et al., 2010] Schulz, J. B., Borkert, J., Wolf, S., Schmitz-Hübsch, T., Rakowicz, M., Mariotti, C., Schoels, L., Timmann, D., van de Warrenburg, B., Dürr, A., Pandolfo, M., Kang,

- J.-S., Mandly, A. G., Nägele, T., Grisoli, M., Boguslawska, R., Bauer, P., Klockgether, T., and Hauser, T.-K. (2010). Visualization, quantification and correlation of brain atrophy with clinical symptoms in spinocerebellar ataxia types 1, 3 and 6. *NeuroImage*, 49(1) :158–168.
- [Schöls et al., 1997] Schöls, L., Amoiridis, G., Büttner, T., Przuntek, H., Epplen, J. T., and Riess, O. (1997). Autosomal dominant cerebellar ataxia : Phenotypic differences in genetically defined subtypes? : Phenotypes in Dominant Ataxias. *Annals of Neurology*, 42(6) :924–932.
- [Schöls et al., 2004] Schöls, L., Bauer, P., Schmidt, T., Schulte, T., and Riess, O. (2004). Autosomal dominant cerebellar ataxias : clinical features, genetics, and pathogenesis. *The Lancet Neurology*, 3(5) :291–304.
- [Seidel et al., 2012] Seidel, K., Siswanto, S., Brunt, E. R. P., den Dunnen, W., Korf, H.-W., and Rüb, U. (2012). Brain pathology of spinocerebellar ataxias. *Acta Neuropathologica*, 124(1) :1–21.
- [Sequeiros and Coutinho, 1993] Sequeiros, J. and Coutinho, P. (1993). Epidemiology and clinical aspects of Machado-Joseph disease. *Advances in Neurology*, 61 :139–153.
- [Shibata-Hamaguchi et al., 2009] Shibata-Hamaguchi, A., Ishida, C., Iwasa, K., and Yamada, M. (2009). Prevalence of Spinocerebellar Degenerations in the Hokuriku District in Japan. *Neuroepidemiology*, 32(3) :176–183.
- [Simon et al., 2007] Simon, D. K., Zheng, K., Velázquez, L., Santos, N., Almaguer, L., Figueroa, K. P., and Pulst, S.-M. (2007). Mitochondrial Complex I Gene Variant Associated With Early Age at Onset in Spinocerebellar Ataxia Type 2. *Archives of Neurology*, 64(7) :1042.
- [Spitzer et al., 1999] Spitzer, R. L., Kroenke, K., and Williams, J. B. (1999). Validation and utility of a self-report version of PRIME-MD : the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. *JAMA*, 282(18) :1737–1744.
- [Sylvestre and Abrahamowicz, 2008] Sylvestre, M.-P. and Abrahamowicz, M. (2008). Comparison of algorithms to generate event times conditional on time-dependent covariates. *Statistics in Medicine*, 27(14) :2618–2634.
- [Synofzik and Ilg, 2014a] Synofzik, M. and Ilg, W. (2014a). Motor Training in Degenerative Spinocerebellar Disease : Ataxia-Specific Improvements by Intensive Physiotherapy and Exergames. *BioMed Research International*, 2014 :1–11.
- [Synofzik and Ilg, 2014b] Synofzik, M. and Ilg, W. (2014b). Motor Training in Degenerative Spinocerebellar Disease : Ataxia-Specific Improvements by Intensive Physiotherapy and Exergames. *BioMed Research International*, 2014 :1–11.
- [Tang et al., 2000] Tang, B., Liu, C., Shen, L., Dai, H., Pan, Q., Jing, L., Ouyang, S., and Xia, J. (2000). Frequency of SCA1, SCA2, SCA3/MJD, SCA6, SCA7, and DRPLA CAG trinucleotide repeat expansion in patients with hereditary spinocerebellar ataxia from Chinese kindreds. *Archives of Neurology*, 57(4) :540–544.
- [Taylor et al., 2005] Taylor, J. M., Yu, M., and Sandler, H. M. (2005). Individualized Predictions of Disease Progression Following Radiation Therapy for Prostate Cancer. *Journal of Clinical Oncology*, 23(4) :816–825.
- [Teive et al., 2008] Teive, H. A. G., Munhoz, R. P., Raskin, S., and Werneck, L. C. (2008). Spinocerebellar ataxia type 6 in Brazil. *Arquivos De Neuro-Psiquiatria*, 66(3B) :691–694.
- [Tezenas du Montcel et al., 2014] Tezenas du Montcel, S., Durr, A., Bauer, P., Figueroa, K. P., Ichikawa, Y., Brussino, A., Forlani, S., Rakowicz, M., Schöls, L., Mariotti, C., van de Warrenburg, B. P., Orsi, L., Giunti, P., Filla, A., Szymanski, S., Klockgether, T., Berciano, J., Pandolfo,

- M., Boesch, S., Melegh, B., Timmann, D., Mandich, P., Camuzat, A., Goto, J., Ashizawa, T., Cazeneuve, C., Tsuji, S., Pulst, S.-M., Brusco, A., Riess, O., Brice, A., and Stevanin, G. (2014). Modulation of the age at onset in spinocerebellar ataxia by CAG tracts in various genes. *Brain*, 137(9) :2444–2455.
- [Thompson et al., 2014] Thompson, A. M., Turner, R. M., Hayen, A., Aniss, A., Jalaty, S., Leary, D. L., Sidhu, S., Delbridge, L., Yeh, M. W., Clifton-Bligh, R., and Sywak, M. (2014). A Preoperative Nomogram for the Prediction of Ipsilateral Central Compartment Lymph Node Metastases in Papillary Thyroid Cancer. *Thyroid*, 24(4) :675–682.
- [Todd and Paulson, 2009] Todd, P. K. and Paulson, H. L. (2009). RNA mediated neurodegeneration in repeat expansion disorders. *Annals of Neurology*, pages NA–NA.
- [Vale et al., 2010] Vale, J., Bugalho, P., Silveira, I., Sequeiros, J., Guimarães, J., and Coutinho, P. (2010). Autosomal dominant cerebellar ataxia : frequency analysis and clinical characterization of 45 families from Portugal. *European Journal of Neurology*, 17(1) :124–128.
- [van de Warrenburg et al., 2002] van de Warrenburg, B. P. C., Sinke, R. J., Verschuuren-Bemelmans, C. C., Scheffer, H., Brunt, E. R., Ippel, P. F., Maat-Kievit, J. A., Dooijes, D., Notermans, N. C., Lindhout, D., Knoers, N. V. a. M., and Kremer, H. P. H. (2002). Spinocerebellar ataxias in the Netherlands : prevalence and age at onset variance analysis. *Neurology*, 58(5) :702–708.
- [van Houwelingen, 2000] van Houwelingen, H. C. (2000). Validation, calibration, revision and combination of prognostic survival models. *Statistics in Medicine*, 19(24) :3401–3415.
- [Verbeke and Molenberghs, 2009a] Verbeke, G. and Molenberghs, G. (2009a). *Linear Mixed Models for Longitudinal Data*. Springer, Dordrecht. OCLC : 851971759.
- [Verbeke and Molenberghs, 2009b] Verbeke, G. and Molenberghs, G. (2009b). *Linear Mixed Models for Longitudinal Data*. Springer, Dordrecht. OCLC : 851971759.
- [Vernerey et al., 2016] Vernerey, D., Huguet, F., Vienot, A., Goldstein, D., Paget-Bailly, S., Van Laethem, J.-L., Glimelius, B., Artru, P., Moore, M. J., André, T., Mineur, L., Chibaudel, B., Benetkiewicz, M., Louvet, C., Hammel, P., and Bonnetain, F. (2016). Prognostic nomogram and score to predict overall survival in locally advanced untreated pancreatic cancer (PROLAP). *British Journal of Cancer*, 115(3) :281–289.
- [Whynes, 2008] Whynes, D. K. (2008). Correspondence between EQ-5d health state classifications and EQ VAS scores. *Health and Quality of Life Outcomes*, 6(1) :94.
- [Williams and Paulson, 2008] Williams, A. J. and Paulson, H. L. (2008). Polyglutamine neurodegeneration : protein misfolding revisited. *Trends in Neurosciences*, 31(10) :521–528.
- [Yamada et al., 2007] Yamada, M., Sato, T., Tsuji, S., and Takahashi, H. (2007). CAG repeat disorder models and human neuropathology : similarities and differences. *Acta Neuropathologica*, 115(1) :71–86.
- [Yasui et al., 2014] Yasui, K., Yabe, I., Yoshida, K., Kanai, K., Arai, K., Ito, M., Onodera, O., Koyano, S., Isozaki, E., Sawai, S., Adachi, Y., Sasaki, H., Kuwabara, S., Hattori, T., Sobue, G., Mizusawa, H., Tsuji, S., Nishizawa, M., and Nakashima, K. (2014). A 3-year cohort study of the natural history of spinocerebellar ataxia type 6 in Japan. *Orphanet Journal of Rare Diseases*, 9(1).