



HAL
open science

Standardisation via CDISC : intérêt et limites pour la gestion de données et la réalisation d'analyses statistiques au sein de l'USMR

Pauline Gaffez

► **To cite this version:**

Pauline Gaffez. Standardisation via CDISC : intérêt et limites pour la gestion de données et la réalisation d'analyses statistiques au sein de l'USMR. Santé publique et épidémiologie. 2019. dumas-02402992

HAL Id: dumas-02402992

<https://dumas.ccsd.cnrs.fr/dumas-02402992>

Submitted on 10 Dec 2019

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.

Master Sciences, Technologies, Santé
Mention Santé Publique

Parcours Systèmes d'Information et Technologies Informatiques pour la Santé

Promotion 2018-2019

**« Standardisation via CDISC : intérêt et limites pour la
gestion de données et la réalisation d'analyses statistiques
au sein de l'USMR. »**

*Mémoire réalisé dans le cadre d'un apprentissage effectué du 10/09/18 au
31/08/19 au sein de l'Unité de Soutien Méthodologique à la Recherche
clinique et épidémiologique (USMR) au CHU de Bordeaux.*

Maître d'apprentissage : **Séverine Martiren**, data manager



Soutenu publiquement le 09/09/19 par **Pauline Gaffez**

Jury de soutenance :

Tutrice universitaire : **Fleur Mouglin**, maître de conférence

Rapporteur : **Vianney Jouhet**, praticien hospitalier

REMERCIEMENTS

Je tiens à remercier :

Ma maître d'apprentissage,

Madame Séverine Martiren,

Data manager, Unité de Soutien Méthodologique à la Recherche clinique et épidémiologique (USMR) du CHU de Bordeaux,

Pour son suivi et son accueil.

Le responsable de l'USMR,

Monsieur Rodolphe Thiébaud,

Pour sa confiance et l'opportunité qu'il m'a offerte.

La responsable du master,

Madame Fleur Mouglin,

Pour son écoute, son accompagnement et sa disponibilité tout au long de ces deux années.

Je tiens également à exprimer ma reconnaissance :

A Monsieur Olivier Quintin,

Pour son temps et ses précieux conseils.

A Madame Emmanuelle Floch-Galaud,

Pour ses conseils bibliographiques.

A toute l'équipe de l'USMR,

Pour leur accueil et leur travail.

Table des matières

1. INTRODUCTION.....	3
2. CONTEXTE.....	4
2.1. Présentation de l'USMR.....	4
2.2. Déroulement d'une étude clinique	5
2.3. La gestion de données biomédicales et les standards.....	5
2.4. Qu'est-ce que le CDISC ?	6
2.4.1. Présentation générale.....	6
2.4.2. Les standards CDISC	7
2.4.3. Zoom sur le CDASH et le SDTM	8
2.4.4. Exemple illustratif de passage au SDTM.....	10
2.4.5. Exemples de mises en place du CDISC	11
2.4.6. CDISC vs. HL7 FHIR	12
2.5. L'étude SOStrial.....	13
2.6. Objectif du stage.....	14
3. METHODES	15
3.1. Mise en place du CDISC sur SOStrial	15
3.1.1. Découverte du CDISC.....	15
3.1.2. Appropriation du CDISC par la définition d'un SDTM	15
3.1.3. Mise en place d'une standardisation s'inspirant du CDISC.....	15
3.1.4. Export des tables	16
3.2. Programmation	18
3.2.1. Définition du cahier des charges	18
3.2.2. Logiciel utilisé.....	18
3.2.3. Etapes du développement.....	18
3.2.4. Optimisation du programme.....	21
3.2.5. Edition d'un rapport et notice explicative	21
3.3. Validation du programme.....	22
3.3.1. Authentification du programme	22
3.3.2. Qualification.....	22
4. RESULTATS	23
4.1. Mise en place du CDISC sur SOStrial	23
4.1.1. SDTM.....	23

4.1.2. DOB	24
4.1.3. Paramétrage des tables d'export.....	25
4.1.4. Tables d'export.....	25
4.2. Programmation	26
4.2.1. Cahier des charges.....	26
4.2.2. Programme(s)	27
4.2.3. Notice explicative.....	32
4.3. Validation du programme.....	34
4.4. Intervention forum méthodologique.....	35
5. DISCUSSION	36
5.1. Faisabilité et intérêt de la mise en place du CDISC au sein de l'USMR	36
5.1.1. Mise en place du CDISC sur SOStrial	36
5.1.2. Implications de la mise en place du CDISC à l'USMR	37
5.1.3. Refonte des circuits de données	39
5.1.4. Intérêt du CDISC pour l'USMR.....	40
5.2. Performances du programme et limites.....	41
5.2.1. Performances	41
5.2.2. Limites liées au CDISC-like.....	41
5.2.3. Autres limites identifiées.....	42
5.2.4. Pistes d'amélioration	42
6. CONCLUSION	44
7. REFERENCES.....	45
8. ANNEXES	47

1. INTRODUCTION

L'Unité de Soutien Méthodologique à la Recherche clinique et épidémiologique (USMR) gère des études cliniques promues par le Centre Hospitalier Universitaire (CHU) de Bordeaux. Dans ce cadre, une partie de ses fonctions est de s'assurer de la bonne gestion des données récoltées lors de ces études. L'équipe est composée de plusieurs gestionnaires de données et il est donc nécessaire que les pratiques soient standardisées afin que toutes les études soient traitées avec les mêmes procédures.

Le *Clinical Data Interchange Standard Consortium* est un consortium international à but non lucratif qui promeut la standardisation des formats de recueil et d'échange des données des études cliniques. Il a publié différents standards qui interviennent chacun à un moment de la recherche, du développement du protocole jusqu'à l'analyse.

Un projet de recherche et développement a été entamé à l'USMR afin de mettre en place un standard CDISC pour une étude prise en charge par l'unité. Le but était d'étudier dans quelle mesure le CDISC est applicable dans le cadre des activités de l'USMR. Le projet a été repris dans le cadre de ce mémoire afin d'étudier la faisabilité et l'intérêt de la mise en place du CDISC pour la gestion des données et les analyses statistiques à l'USMR. Seul le *Study Data Tabulation Model* (SDTM) a été appliqué dans ce cadre et, afin de ne pas bloquer les analyses statistiques pour lesquelles des programmes standards sont utilisés, le choix a été fait de développer un programme informatique afin de transposer automatiquement les tables pour les rendre à nouveau exploitables.

Après avoir présenté le contexte de ce travail, ce mémoire explique les méthodes employées pour mettre en place le SDTM sur l'étude pilote et pour développer le programme. Les résultats sont ensuite présentés et des discussions sont ouvertes à propos des standards et plus précisément de l'intérêt du CDISC pour les pratiques actuelles de l'USMR. Les performances et les limites du programme sont également discutées.

2. CONTEXTE

2.1. Présentation de l'USMR

Depuis le 1^{er} mai 1999, l'Unité de Soutien Méthodologique à la Recherche clinique et épidémiologique (USMR) fait partie du CHU de Bordeaux et plus particulièrement du Service d'Information Médicale (SIM) du pôle de santé publique (1). L'unité est aujourd'hui dirigée par Rodolphe Thiébaud et par son adjoint, Antoine Bénard. L'USMR traite les études de recherche clinique ou épidémiologique dont le CHU de Bordeaux est le promoteur. Pour répondre aux demandes de plus en plus nombreuses et de plus en plus complexes, elle a développé des compétences scientifiques et techniques. Elle intervient aujourd'hui sur les plans épidémiologique grâce aux méthodologistes, biostatistique grâce aux statisticiens et informatique grâce aux gestionnaires de données biomédicales et à l'informaticien de l'équipe. Les projets reçus au sein de l'unité sont étudiés et leur validité scientifique est évaluée avant de pouvoir prétendre à un financement pour permettre une mise en place la plus rapide possible. Les missions de l'USMR sont triples (1) :

- elle apporte une aide méthodologique pour la constitution et le suivi des protocoles
- elle prend en charge tout ou partie de la logistique des études
- elle participe à la valorisation des résultats des recherches.

Depuis 2014, l'USMR est certifiée sur la norme ISO 9001 (1). Cela fait suite à la mise en place d'un système de management de la qualité (SMQ) en vue de satisfaire les exigences réglementaires françaises et européennes ainsi que les bonnes pratiques de recherche clinique. Cette politique de qualité est sans cesse revue et modifiée afin d'améliorer l'efficacité du travail de l'équipe et de pouvoir gérer les risques et les opportunités qui pourraient être rencontrés. Les trois axes de la politique qualité et les objectifs associés pour 2016-2020 sont (1) :

- Un haut niveau de professionnalisme, avec pour objectif d'optimiser la mise en place, le suivi et la valorisation des études cliniques.
- Une maîtrise et une amélioration continue des processus, avec pour objectif d'optimiser la gestion des données des études cliniques.
- Une adaptation des compétences à l'évolution des exigences scientifiques, avec pour objectif de faciliter la formation aux méthodes statistiques et épidémiologiques actuelles.

C'est dans le deuxième axe que s'inscrit la mission de stage qui m'a été confiée. Elle fait suite au besoin d'optimiser le temps nécessaire au passage des données brutes saisies sur le cahier d'observation électronique (eCRF pour *electronic Case Report Form*) en tables exploitables par les statisticiens pour les analyses. Pour ce faire, l'intérêt de la standardisation a été étudié par le biais des standards du CDISC (*Clinical Data Interchange Standard Consortium*) et plus particulièrement du SDTM (*Standard Data Tabulation Model*).

2.2. Déroulement d'une étude clinique

Les Entreprises du Médicament (LEEM) définissent une étude clinique comme « une situation expérimentale au cours de laquelle on teste chez l'homme la véracité ou non d'une hypothèse » (2). Dans la recherche clinique, toute étude commence par la rédaction d'un protocole. Il s'agit d'un document qui justifie et explicite la question de recherche à l'aide d'une revue de la littérature sur le sujet ; cette justification comprend également les bénéfices attendus ou potentiels de la recherche pour les patients (2). Le ou les objectifs sont ensuite défini(s) précisément avec les critères de jugement (critères sur lesquels se baseront les conclusions de l'étude) (2). Ce document est relu par différents experts, notamment à l'USMR le méthodologiste et le statisticien. Le protocole est donc un document qui explicite les différents aspects du déroulement de l'étude. Il inclut donc notamment (2) :

- La justification de l'étude
- Des considérations éthiques
- Les objectifs et critères de mesure
- La sélection de la population de l'étude
- Le plan expérimental et la description éventuelle des traitements
- Les méthodes statistiques et la gestion des données
- Le lieu et la durée de la recherche
- Les procédures à suivre et les réglementations à respecter

2.3. La gestion de données biomédicales et les standards

Une fois le protocole rédigé et validé par les autorités compétentes (Commission Nationale de l'Informatique et des Libertés (CNIL), Comité de Protection des Personnes (CPP), etc.), les études cliniques se déroulent la plupart du temps en trois phases : la collecte des données à l'aide d'un cahier d'observation, l'utilisation de ce cahier pour créer la base de données et enfin la production d'un rapport d'analyse (3). Il en résulte que la qualité de l'analyse dépend des étapes qui la précèdent et donc de la qualité de la base de données (3).

L'objectif principal de la gestion de données cliniques est de fournir des données de qualité pour les analyses, c'est-à-dire des données avec le moins d'incohérences et le moins de données manquantes possible. Pour parvenir à cet objectif, des bonnes pratiques sont définies afin d'assurer la complétude des données, leur fiabilité et la validité des traitements effectués (4). La gestion de données biomédicales a évolué en réponse à la forte demande d'accélérer les processus de mise en place d'études cliniques et au développement des nouvelles technologies, passant de cahiers d'observation papier à des cahiers électroniques. Ces nouvelles pratiques nécessitent, pour améliorer la qualité et gagner en efficacité, de standardiser les procédures et les terminologies mises en œuvre (4).

Selon l'Organisation de Coopération et de Développement Economiques (OCDE), la standardisation est « le processus de parvenir à un accord sur les définitions communes des données, les formats, la représentation et les structures de toutes les couches de données et éléments » (5). Selon l'ISO, les standards sont « des accords documentés contenant des

spécifications techniques ou d'autres critères précis pouvant être utilisés systématiquement, comme des règles, des recommandations, ou des définitions de caractéristiques pour assurer que le matériel, les produits, les procédures, et les services correspondent à l'objectif » (6).

Dans la recherche, les standards fournissent une connaissance commune qui définit non seulement comment échanger les données mais aussi comment les collecter, les comprendre et les utiliser dans un but commun : mieux soigner les patients (6). Ils garantissent une meilleure qualité des données grâce à la cohérence qu'ils fournissent. Il existe différents types et niveaux de standards : il peut s'agir de terminologies, de règles définissant les informations à récolter dans un cahier d'observation, de structures pour les tables d'analyses statistiques ou encore de métadonnées (6).

En pratique, le recours à des standards permet de faciliter les échanges d'informations et l'intégration de différentes sources de données dans des études, des registres ou des entrepôts de données par exemple (7). Il permet également de faciliter les méta-analyses en simplifiant l'agrégation des données provenant de diverses études précédemment menées. Enfin, il permet parfois la soumission à différentes autorités telles que la *Food and Drug Administration* (FDA) aux Etats-Unis ou la *Pharmaceuticals and Medical Devices Agency* (PMDA) au Japon (7).

Il existe aujourd'hui des standards internationaux permettant de définir des règles de collecte et d'organisation de données comme les standards CDISC ou HL7 FHIR (*Fast Healthcare Interoperability Ressources*). Cela permet d'optimiser la gestion de données par le recours à une structure unique et donc de faciliter l'interopérabilité, l'interopérabilité désignant « la capacité de systèmes hétérogènes à échanger leurs données, de sorte que celles émises par l'un puissent être reconnues et interprétées par les autres, utilisées et traitées » (8). Cela va dans le sens des principes FAIR (*Findable, Accessible, Interoperable and Reusable*), publiés en 2016 dans la revue *Scientific Data* et requis par la Commission Européenne pour la recherche scientifique (9). Ces principes non spécifiques à la recherche clinique visent à promouvoir le partage de données pour permettre leur réutilisation par les humains et les machines. Le but étant à terme, comme le précise l'ICMJE (*International Committee of Medical Journal Editors*) dans un article publié en 2017, de « créer un environnement dans lequel le partage de données désidentifiées devient la norme »(10). Il est en effet important de développer le partage de données non seulement dans un souci de vérification des résultats publiés, mais également car cela permet de creuser de nouvelles questions de recherche en réutilisant des données existantes, donc sans avoir à recueillir à nouveau des données ce qui permet un gain en termes d'efficience (10).

2.4. Qu'est-ce que le CDISC ?

2.4.1. Présentation générale

« Le CDISC (*Clinical Data Interchange Standards Consortium*) est un consortium international à but non lucratif qui vise à promouvoir la standardisation des formats de recueil, d'échange, de soumission et d'archivage de données dans la recherche clinique » (11). Le but étant de fluidifier la circulation des données depuis leur récolte jusqu'à leur analyse afin de réduire les temps de

traitement et donc les coûts. Pour ce faire, CDISC a créé et publié une série de standards présentés dans la Figure 1 qui sont régulièrement mis à jour pour (11) :

- Modéliser / Organiser
- Recueillir
- Echanger (et archiver)
- Soumettre les données aux autorités de santé (pour l'obtention d'une autorisation de mise sur le marché par exemple)
- Présenter l'analyse des données à ces mêmes autorités

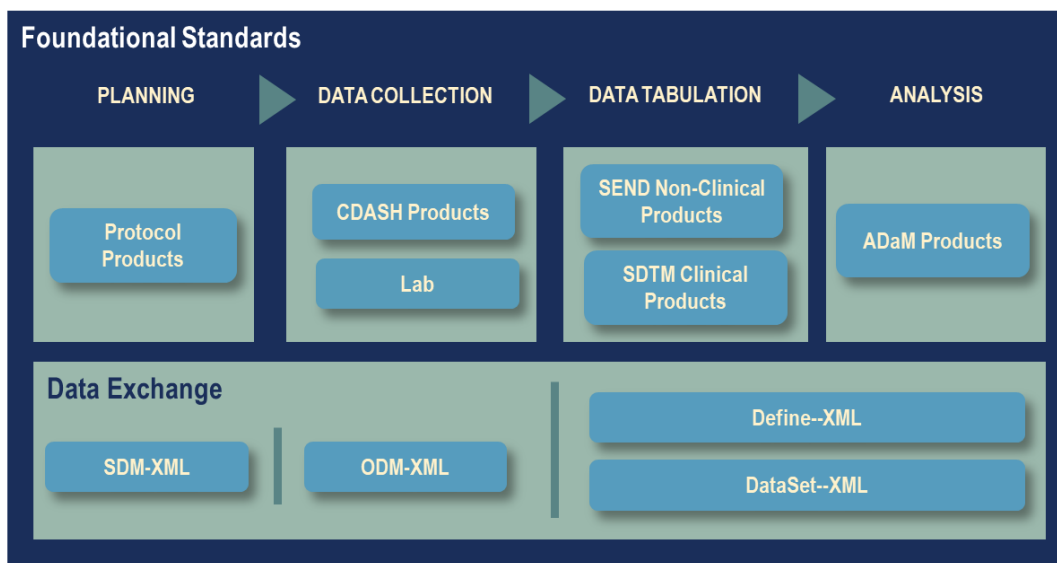


Figure 1 : ODM-based standards supporting the CDISC foundational standards content (12).

Selon CDISC, l'adoption de ces standards permet (11) :

- Une harmonisation des formats de variables et des outils de recueil
- La facilitation des échanges de données entre différents systèmes d'information (par exemple entre Dossier Patient Electronique et eCRF)
- Un archivage des eCRF à long terme, dans un format ouvert xml (Extensible Markup Language)
- Une soumission des données de la recherche aux autorités de santé dans des formats stricts, leur permettant de retrouver par elles-mêmes les résultats amenés à être publiés (obligatoire pour la FDA aux USA dès 2017)
- D'avantage d'indépendance vis-à-vis des éditeurs de logiciels
- Etc.

2.4.2. Les standards CDISC

Il existe donc différents standards développés par CDISC, intervenant à différentes étapes de la recherche, de sa planification jusqu'à l'analyse mais également l'archivage et la soumission aux autorités. Voici une brève description de chacun d'eux (11) :

- **BRIDG** (*Biomedical Research Integrated Domain Group*) : ce standard permet de cartographier le protocole sous la forme d'une série de diagrammes UML (*Unified*

Modeling Language) afin de visualiser les différentes étapes et flux de données de l'étude.

- PRM (*Protocol Representation Model*) : il s'agit d'un « outil d'aide à la conception du protocole et à la création des documents. »
- CDASH (*Clinical Data Acquisition Standard Harmonization*) : ce standard est présenté dans la partie suivante (section 2.4.3.1, page 8), il permet de standardiser les eCRF. Il existe des *mappings* entre le CDASH et le SDTM.
- LAB (*Laboratory Standards*) : ce standard vise à faciliter l'intégration des données issues des laboratoires d'analyse.
- ODM (*Operational Data Model*) : il s'agit d'un standard pour « l'acquisition, l'échange et l'archivage des données cliniques » qui a recours au format XML. Un fichier ODM contient à la fois les données et les métadonnées (nécessaires à la définition et à la restitution des données) de l'étude.
- SDM (*Study Design Model*) : ce n'est pas un standard en soit mais une « série de fragments d'information, d'arborescences XML, que l'on peut rajouter dans un fichier ODM. »
- SDTM (*Study Data Tabulation Model*) : ce standard est présenté en détail dans la partie suivante (section 2.4.3.2, page 9). Il permet de définir les tables de données de l'étude.
- SEND (*Standard for Exchange of Non-clinical Data*) : ce standard est une adaptation du SDTM pour des données non-cliniques (le sujet n'est pas l'humain mais l'animal).
- ADaM (*Analysis Dataset Model*) : ce standard permet la « soumission des données d'analyse statistique aux organismes réglementaires comme la FDA. »
- SHARE (*Shared Health and Research Electronic Library*) : il s'agit d'un « projet de bibliothèque électronique de métadonnées au format CDISC. »

2.4.3. Zoom sur le CDASH et le SDTM

L'application de CDISC dans le cadre de ce stage prend place au sein de la gestion des données et concerne donc plus précisément les standards CDASH et SDTM, et en particulier ce dernier qui permet la structuration des tables.

2.4.3.1. Présentation du CDASH

« Le *Clinical Data Acquisition Standards Harmonization* (CDASH) a pour objectif de standardiser le recueil des données sur les CRF électroniques ou papiers. [...] Le but est de réduire la variabilité entre les CRF, en particulier en réutilisant les mêmes définitions de variables » (13). Il s'agit d'un guide de bonnes pratiques qui suggère quelles données recueillir et comment le faire au mieux pour faciliter leur traitement. Le CDASH a été développé de manière pragmatique par l'*Association of Clinical Research Organisation* (ACRO) qui a analysé les variables récoltées dans les cahiers des clients (13). Le lien a ensuite été fait avec les exigences de soumission des données aux autorités grâce au standard SDTM (explicité ci-

après). Le CDASH est utilisé pour identifier et lister les données à recueillir, tandis que le SDTM entre plus dans les détails, notamment au niveau des métadonnées (13).

2.4.3.2. Présentation du SDTM

« Le *Study Data Tabulation Model* (SDTM) est un modèle conceptuel qui permet la présentation standardisée des données des études cliniques [...] dans le but de les soumettre à des autorités réglementaires telles que la FDA » (14). Il a été développé par une équipe CDISC et permet la structuration des bases de données sous la forme de tables (appelées domaines) au format SAS transport ou XML. Pour pouvoir être soumises aux autorités, les tables doivent être supplémentées d'un fichier XML appelé DEFINE qui contient les métadonnées de la structure et le format des données transmises (14). Le CDISC décrit le SDTM dans deux documents fondamentaux (téléchargeables sur leur site internet¹ après création d'un compte utilisateur) :

- le *Study Data Tabulation Model* qui décrit le modèle lui-même
- le *SDTM implementation guide* qui a pour objectif de guider les utilisateurs dans l'implémentation du modèle et de fournir des informations complémentaires et des exemples.

L'un des concepts fondamentaux du SDTM est le concept de domaine. Un domaine peut être décrit comme un regroupement de données concernant un même thème. Les domaines se regroupent eux-mêmes en plusieurs classes d'observations. Les données d'une étude peuvent majoritairement être classées en trois classes d'observations (14) :

- Les interventions (*Interventions*) : ensemble des observations relatives aux traitements des sujets :
 - Traitements spécifiés par le protocole de l'étude (ex. : médicament de l'étude)
 - Traitements administrés durant l'étude (ex. : traitements concomitants)
 - Traitements auto-administrés par les sujets (ex. : caféine, tabac, alcool)
- Les événements (*Events*) : ensemble des événements relatifs à l'étude :
 - Événements prévus par le protocole (ex. : randomisation)
 - Événements non planifiés prenant place durant l'étude (ex. : effets indésirables)
 - Événements non planifiés ayant eu lieu auparavant (ex. : historique médical)
- Les résultats (*Findings*) : ensemble des observations résultant de tests et d'évaluations prévues par le protocole telles que le poids, la taille, les données de laboratoire, les réponses aux questionnaires, etc.

Les données ne faisant partie d'aucune de ces trois catégories seront dans des tables avec des contenus spécifiques comme les informations relatives aux sujets, au schéma de l'étude ou aux relations entre les données (14). La liste exhaustive des domaines développés par le CDISC a été mise en annexe (ANNEXE 1: Domaines du SDTM, page 48). Pour répondre aux besoins, CDISC crée régulièrement de nouveaux domaines. Cependant, si un domaine est nécessaire à

¹ Consultable à l'adresse www.cdisc.org

une étude et qu’il n’existe pas encore, il est possible de le créer puis de le soumettre au groupe de travail CDISC-SDTM. Le CDISC souligne que « ce processus doit suivre des modalités précises tant au niveau de la création du domaine (définition et attribution des variables) que du nom qui lui est donné »(14).

Le deuxième concept central du SDTM est le concept d’observation. Il s’agit d’une « série d’informations recueillies au cours d’une étude clinique » (14). Une ligne dans une table de données définit donc une observation. Une observation est décrite par plusieurs variables (sujet, date, question, réponse, unité, etc.) pouvant être classées selon leur rôle. Par exemple le numéro du sujet a un rôle d’identification. De plus, les noms des variables du SDTM répondent à un certain nombre de conventions. Les noms des variables sont précédés des deux lettres servant à identifier le domaine auquel elles appartiennent (exemple : **A**ETERM, **M**HTERM, **L**BTEST) excepté pour les variables appartenant à plusieurs domaines (ex. : STUDYID, USUBJID, VISIT, VISITNUM) (14).

2.4.4. Exemple illustratif de passage au SDTM

Ce qu’il faut bien comprendre dans le SDTM est le passage d’une structure horizontale (pratique actuelle de l’USMR), c’est-à-dire qu’une ligne correspond à l’ensemble des observations pour un patient, à une structure verticale dans laquelle une ligne correspond à une seule observation (contenant plusieurs informations telles que la mesure, l’unité et la date par exemple) d’un patient. À noter tout de même que certains domaines (AE, DM, MH, CM entre autres) gardent une structure horizontale. Il s’agit des domaines les plus courants des études cliniques (événements indésirables, données démographiques, antécédents et traitements), soumis à peu de variabilité et donc très rigides dans le choix des informations à renseigner. Les autres domaines ont une structure souple grâce aux variables –TEST (qui enregistre la question posée) et –ORRES (qui enregistre la réponse à la question) qui permettent de définir toutes les observations souhaitées. Les Tableau 1 et Tableau 2 proposent un exemple de table contenant des observations cliniques (domaine « Vital Signs ») avant et après verticalisation.

Tableau 1 : Exemple classique de table horizontale

IDSTUDY	IDPAT	VISDAT	POIDS	POIDS_U	TAILLE	TAILLE_U	EXAMCLIN
EXTRIAL	1234	12/01/2018	3.6	Kg	51	Cm	Normal
EXTRIAL	1234	26/05/2018	5.9	Kg	58	Cm	Normal
EXTRIAL	5678	19/09/2018	7.3	Kg	64	Cm	Normal

Tableau 2 : Exemple de table après verticalisation

STUDYID	DOMAIN	USUBJID	VSTEST	VSTESTCD	VSORRES	VSORRESU	VSUTC
EXTRIAL	VS	1234	Poids	PDS	3.6	Kg	12/01/2018
EXTRIAL	VS	1234	Taille	TAILLE	51	Cm	12/01/2018

STUDYID	DOMAIN	USUBJID	VSTEST	VSTESTCD	VSORRES	VSORRESU	VSDTC
EXTRIAL	VS	1234	Examen clinique	EC	Normal		12/01/2018
EXTRIAL	VS	1234	Poids	PDS	5.9	Kg	26/05/2018
EXTRIAL	VS	1234	Taille	TAILLE	58	Cm	26/05/2018
EXTRIAL	VS	1234	Examen clinique	EC	Normal		12/01/2018
EXTRIAL	VS	5678	Poids	PDS	7.3	Kg	19/09/2018
EXTRIAL	VS	5678	Taille	TAILLE	64	Cm	19/09/2018
EXTRIAL	VS	5678	Examen clinique	EC	Normal		19/09/2018

2.4.5. Exemples de mises en place du CDISC

2.4.5.1. Les *Therapeutic Area User Guides* (TAUG)

En partenariat avec divers acteurs tels que des institutions ou des associations, CDISC a créé des guides définissant des normes pour permettre la mise en place des standards (en particulier le SDTM et les terminologies de référence) dans différentes aires thérapeutiques (15). Il existe à ce jour 31 guides publiés, notamment pour Alzheimer, le cancer du sein, le diabète, l'hépatite C, la maladie de Parkinson ou la schizophrénie (15). Par exemple le guide pour la mise en place du CDISC dans le cadre des données du paludisme a été créé en collaboration avec le *WorldWide Antimalarial Resistance Network* (16). Ces guides permettent de standardiser le recueil de données dans un champ thérapeutique précis, ce qui permet ensuite de pouvoir agréger des données provenant de diverses sources et de pouvoir les explorer et gagner en puissance de calcul (15).

Toujours sur le site internet² du CDISC, des articles scientifiques sont mis à disposition faisant état de cas d'usage dans différents domaines, à la fois académiques et industriels (17).

2.4.5.2. Quelques exemples d'utilisation du CDISC

Différents articles dans la littérature font la preuve de l'utilité du CDISC dans le cadre de la standardisation à des fins d'interopérabilité. Ainsi, en 2014 Matsumara et son équipe (18) sont parvenus à connecter des eCRF à des systèmes d'enregistrements médicaux électroniques à l'aide du standard ODM. Cela a permis de remplir automatiquement les eCRF et donc de ne pas avoir à transcrire manuellement les données lues dans un dossier médical ou sur un cahier d'observation papier. Cette piste d'exploration est très prometteuse étant donnée la tendance de la recherche clinique à recourir aux données des soins (18).

Un autre exemple a été publié en 2018 : au centre Antoine Lacassagne à Nice, un entrepôt a été créé pour stocker les données des essais thérapeutiques (19). Au total, 16 études de cancérologie

² Consultable à l'adresse www.cdisc.org

qui avaient été promues par le centre et qui possédaient chacune une base de données électronique ont pu être agrégées en une seule et même base de données. Ainsi, les statisticiens disposent d'une « grande base de données fiable, requêtable et analysable » (19). Cette étude de faisabilité permet de dire que la création d'un entrepôt de données plus complet est possible.

Cela vient corroborer l'action mise en place par le Data Center UNICANCER R&D depuis 2011 (20). A l'aide du SDTM, l'équipe a pu définir les domaines à utiliser et ceux à créer plus spécifiquement pour l'oncologie. Ils ont ainsi pu définir des CRF-types en fonction des contraintes de l'étude. Pour chaque CRF-type, des tables décrivant les métadonnées ont ensuite pu être créées pour permettre de faire le lien entre le CRF annoté et l'architecture de la base de données. Enfin, des listes de codes issues des terminologies CDISC ont été dressées. Tout cela a permis de créer une structure reproductible qui « permet une exploitation optimale des bases et une simplification de la mise en commun des données des essais pour une exploitation méta-analytique optimisée »(20).

A plus large échelle, le *National Center Biobank Network* (NCBN) a publié en 2016 une étude de faisabilité concernant la standardisation du recueil de données de six biobanques nationales à l'aide du CDISC (21). Bien que l'étude conclue qu'il leur faudrait étudier davantage les besoins liés à la standardisation, elle montre que la mise en place du CDISC est réalisable dans la mesure où les 202 items recueillis par le NCBN peuvent être mis en correspondance avec des items du SDTM. L'article précise que 50 à 70% des items requis dans les domaines du SDTM sont alors remplis (21).

2.4.6. CDISC vs. HL7 FHIR

Depuis plus de quinze ans, le CDISC développe des standards destinés à la recherche clinique. Mais un autre standard se développe depuis quelques années, en particulier dans le domaine du soin : il s'agit du standard HL7 FHIR (*Fast Healthcare Interoperability Resources*). HL7 FHIR est un standard facile à appliquer et souple fonctionnant sur le principe de ressources (22). Ces ressources couvrent 80% des données récoltées, les 20% restant trouvent leur place dans les extensions. Une API RESTful existe et il est possible d'utiliser des formats XML, JSON ou Turtle (22).

La recherche clinique tendant de plus en plus à reposer sur les données du soin, le besoin est donc grandissant d'avoir un standard qui recouvre à la fois le soin et la recherche clinique (23). C'est pourquoi la possibilité de fusionner CDISC et HL7 FHIR a été étudiée dans la littérature. Leroux a cherché à créer des *mappings* entre CDISC-ODM et HL7 FHIR (23). Il a en partie pu le faire, mais des liens restent manquants entre les deux standards. Il propose donc l'ajout de certaines ressources pour pouvoir pallier à ces manques. Aerts s'est penché sur le standard ODM v2 qui est en train d'être élaboré et qui pourrait inclure les possibilités développées par HL7 qui ne l'étaient pas par CDISC, comme la diversité de formats (non plus seulement XML mais aussi Turtle ou JSON) (24). Zopf, Abolafia et Reddy ont quant à eux mené un projet pilote consistant à récupérer des données électroniques de soin directement pour la recherche clinique

en étudiant la possibilité de passer aux standards CDISC à partir de FHIR pour pouvoir soumettre un médicament aux autorités (25).

Tous ces articles s'accordent sur le fait qu'il est nécessaire que ces deux standards convergent dans un futur proche, car le fait que les données du soin puissent servir à des fins de recherche est de plus en plus probable et pourrait devenir une grande partie de la recherche à l'avenir. Il est évident cependant que pour le moment en France, cela ne pourrait être applicable compte tenu de la réglementation en vigueur. Mais il est important d'envisager cette possible évolution et de tout faire pour faciliter l'interopérabilité entre les soins et la recherche.

2.5. L'étude SOSrial

L'étude SOSrial est une étude promue par le CHU de Bordeaux dans le cadre d'un programme hospitalier de recherche clinique national (PHRCN) dont l'investigateur principal est le Professeur Jean-François Dartigues. Son protocole³ définit l'objectif principal comme la comparaison de « deux stratégies thérapeutiques pour le risque à long terme (deux ans) de dépendance lourde, institutionnalisation et décès chez des patients présentant une maladie d'Alzheimer légère à modérée avec non-réponse aux IAC (inhibiteurs de l'acétylcholinestérase) à 6 mois : poursuite ou arrêt des IAC ».

La stratégie de cette étude est de recruter les patients atteints d'Alzheimer dans les Centres Mémoire de Ressources et de Recherche (CMRR) et les Centres Mémoire (CM) de France. « L'étude sera proposée à tous les nouveaux malades lors de la première prescription d'IAC. Après six mois de traitement par les IAC, les sujets seront classés selon l'évolution du score au Mini Mental State (MMS) en « non-répondeurs » (perte d'au moins un point de MMS) ou « répondeurs » (MMS inchangé ou augmenté).

- Les répondeurs continueront le traitement par les IAC selon les habitudes du clinicien.
- Les non-répondeurs seront inclus dans l'essai thérapeutique avec une randomisation individuelle en deux groupes : un groupe arrêtant le traitement par les IAC, un groupe les poursuivant.

Tous les patients randomisés seront ensuite suivis pendant deux ans avec recueil des critères de jugement tous les six mois. »

Le critère de jugement principal est un critère combiné associant la dépendance totale à la toilette et/ou à l'habillement et/ou l'entrée en institution et/ou le décès deux ans après la randomisation.

Cette étude se justifie dans un contexte où les IAC sont « les seuls médicaments avec une efficacité reconnue au stade léger à modéré de la maladie [d'Alzheimer] ». Cependant, « leur efficacité est faible à modérée selon les avis et leur remboursement remis en question régulièrement, notamment en raison de l'absence de preuves d'efficacité à long terme sur la dépendance et l'entrée en institution. En 2011, la HAS (Haute Autorité de Santé) a considéré

³ Protocole de l'étude SOSrial version 2.0 du 26/01/2018 (document interne à la structure).

que l'objectif du traitement par les IAC était l'amélioration ou la stabilisation d'une échelle cognitive globale, le MMS après six mois de traitement. La poursuite du traitement est dans ce cas justifiée. Par contre en cas de non réponse au traitement par les IAC c'est-à-dire en cas de baisse du MMS à six mois, l'intérêt de la poursuite des IAC n'est pas évident. » Or, depuis la rédaction du protocole les IAC ont été déremboursés par l'Assurance Maladie sur une décision de la HAS. Cela a mis un coup d'arrêt à l'étude. Des discussions sont en cours sur une possibilité de reprise grâce à un financement des médicaments dans le cadre de l'étude, avec l'ajout potentiel d'un suivi des répondants qui pourrait intéresser la HAS.

2.6. Objectif du stage

Olivier Quintin, data manager (DM) au sein de l'USMR, avait entamé avec d'autres personnes de l'équipe un projet de recherche et développement pour mettre en place le CDISC sur l'étude SOStrial, dont il était le DM avant que cette dernière ne soit arrêtée. Cette démarche prenait racine dans le fait que dans un précédent poste, il avait eu à travailler en CDISC et pensait que cela pouvait être bénéfique à l'USMR. Il souhaitait donc voir dans quelles mesures cela était applicable dans l'unité. L'étude étant prévue sous Ennov Clinical (logiciel de gestion de données utilisé à l'USMR), un eCRF avait été développé sous ce logiciel à l'aide du cahier d'observation validé. Puis un document d'organisation de base de données (DOB) avait été construit, ce document servant de référence au sein de l'USMR pour la création de la base de données. Ce DOB avait permis de commencer à créer les tables d'export au format CDISC sur CSEXport, le module d'export du logiciel Ennov Clinical.

Le groupe des DM avait par ailleurs déjà entamé des efforts de standardisation, notamment avec la création d'un DOB-type. Ce document permet de définir les tables récurrentes au sein des études (« pré-inclusion », « examen clinique », « événements indésirables », « fin d'étude », etc.) et les variables communément recueillies avec une standardisation de leur nom (par exemple « PDS » pour le poids, « CII » pour le premier critère d'inclusion, etc.).

L'objectif du stage est donc de poursuivre ces efforts et de reprendre le projet de recherche et développement, le sous-effectif du groupe de DM de l'USMR ne permettant pas à Olivier Quintin de se concentrer dessus. En reprenant depuis le départ, c'est-à-dire depuis le cahier d'observation qui avait été validé, il s'agit de voir dans quelles mesures le CDISC est applicable à l'étude SOStrial et s'il l'est, dans quelles mesures et avec quels impacts il pourrait être généralisé à toutes les études traitées par l'USMR pour la gestion de données et les analyses. Le but est également de voir s'il faut appliquer formellement le CDISC ou s'il faut uniquement s'en inspirer, dans un souci de standardisation des pratiques de gestion de données au sein de l'USMR. Un cahier des charges doit également être mis en place en collaboration avec les statisticiens. Ce cahier des charges définira leurs attentes vis-à-vis des tables fournies pour analyse et servira de base au travail de programmation nécessaire à l'exploitation des tables par les programmes standards. Le programme développé servira à transposer les tables CDISC verticale en tables horizontales.

3. METHODES

3.1. Mise en place du CDISC sur SOStrial

3.1.1. Découverte du CDISC

Les standards développés par CDISC sont présentés sur le site internet du consortium⁴ et les guides d'implémentation y sont téléchargeables sous réserve de la création d'un compte. Les deux documents fondateurs du SDTM (la description du modèle et le guide d'implémentation) sont donc téléchargeables librement. Des exemples de recours au CDISC sont également décrits. En revanche, la littérature ne fait pas ou très peu de retours d'expérience sur la mise en place du CDISC au sein d'une structure, qui plus est d'une structure traitant de thématiques variées.

Afin de faciliter le travail d'implémentation du SDTM, une traduction est proposée par le groupe des Data Managers Académiques (acaDM). Ce document téléchargeable sur leur site⁵ (26) propose une traduction des domaines et des variables du CDASH et du SDTM.

3.1.2. Appropriation du CDISC par la définition d'un SDTM

Un premier SDTM a été dressé en aveugle du travail effectué par Olivier Quintin. Ce SDTM a pris appui sur le CO rédigé par l'équipe projet et validé par l'USMR. Toutes les observations relevées dans le CO ont été attribuées à un domaine et les variables à garder ont été définies. Les règles du standard ont été respectées autant que faire se peut, dans la limite de leur compréhension et de leur difficulté de mise en place. La discussion engendrée par ce travail préliminaire a permis de faire émerger des questionnements dont il sera question dans la discussion.

3.1.3. Mise en place d'une standardisation s'inspirant du CDISC

Au terme des discussions le choix a été fait de davantage s'inspirer du CDISC que de l'appliquer purement et simplement, pour des raisons qui seront abordées dans la discussion. Des compromis ont été faits entre le travail effectué par Olivier Quintin et le SDTM précédemment dressé. Le DOB existant pour l'étude a été mis à jour. Les compromis effectués sont les suivants :

- Pour éviter la multiplication des tables, des observations normalement récoltées dans différents domaines ont été regroupées. Par exemple le domaine « Subject Characteristics » contient les données démographiques qui se trouvent normalement dans le domaine « Demographics ».
- Les variables du type « --REASND » pour « raison pour laquelle cela n'a pas été fait » ont été remplacées par des variables « --TEXT » au format texte qui peuvent contenir

⁴ Consultable à l'adresse www.cdisc.org

⁵ Consultable à l'adresse www.acadm.fr

n’importe quel type de précision (par exemple lorsque la réponse à une question est « Autre, préciser »).

- Pour une gestion facilitée des formats lors de la transposition future des tables, une variable « –FMT » a également été ajoutée pour récolter les formats des observations en ayant un.
- En plus de la variable « –ORRES » qui contient le résultat original au format numérique, une variable « –STRESC » permet de récupérer le résultat décodé lorsque ce dernier est associé à une bibliothèque. Cette variable sert uniquement de vérification au DM pour le paramétrage des tables d’export.

3.1.4. Export des tables

3.1.4.1. Création des tables d’export

Les tables d’export ont été créées à l’aide du module CSExport d’Ennov Clinical. Chacune des tables à exporter (une table correspond à un domaine) est créée puis définie. Pour chacune des tables les variables sont définies, nommer et leur format est précisé. Ensuite, les champs de l’eCRF correspondant aux données doivent être renseignés pour chacune des variables. Il est possible également d’appliquer certaines formules (définies dans un guide utilisateur), par exemple pour récupérer automatiquement le numéro du patient ou bien pour décoder une valeur associée à une bibliothèque. Au final, la quasi-totalité des champs de l’eCRF doit être retrouvée dans les tables d’export. La Figure 2 donne un aperçu du résultat une fois que la table VS a été paramétrée (certaines variables ont été supprimées dans cet exemple, dans un souci de lisibilité) :

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
NOM VARIABLE	STUDYID	USUBJID	VISIT	VISITDY	VSTEST	VSTESTCD	VSORRES	VSORRESU	VSTEXT	VSFMT
INITITULE	Nom de l'étude	Code Patient	Nom de la visite	Date de la visite	Libellé du test	Code du test	Résultat numérique	Unité	Précision	Format de VSORRES
FORMAT INPUT	\$75.	\$254.	\$254.	\$10.	\$200.	\$10.	8.	\$10.	\$200.	\$10.
LONG. MAX CS	8	254	254	10	44	6	8	4	39	7
BIBLIO										
DATE - HEURE				DDMMYY10./VISITDY1						
OCCURRENCE 1	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.12020)	q3020	q12723	q12725	q12020	q12724		
OCCURRENCE 2	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.12020)	q3020	q12733	q12735	q12030	q12734		
OCCURRENCE 3	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.12020)	q3020	q12743	q12745	q12040	q12744		
OCCURRENCE 4	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.12020)	q3020	q12753	q12755	q12050	q12754		
OCCURRENCE 5	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.12020)	q3020	q12763	q12765	q12060	q12764		
OCCURRENCE 6	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.12020)	q3020	q12773	q12775	q12070		q12090	F01020_
OCCURRENCE 7	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.12020)	q3020	q12783	q12785	q12080		q12090	F01020_
OCCURRENCE 8	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.12020)	q3020	q12803	q12795	q12100			OUIINON
OCCURRENCE 9	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.20020)	q19020	q20723	q20725	q20020	q20724		
OCCURRENCE 10	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.20020)	q19020	q20733	q20735	q20030	q20734		
OCCURRENCE 11	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.20020)	q19020	q20743	q20745	q20040	q20744		
OCCURRENCE 12	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.20020)	q19020	q20753	q20755	q20050	q20754		
OCCURRENCE 13	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.20020)	q19020	q20763	q20765	q20060	q20764		
OCCURRENCE 14	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.20020)	q19020	q20773	q20775	q20070		q20090	F01020_
OCCURRENCE 15	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.20020)	q19020	q20783	q20785	q20080		q20090	F01020_
OCCURRENCE 16	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.20020)	q19020	q20803	q20795	q20100			OUIINON
OCCURRENCE 17	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.31020)	q29020	q31723	q31725	q31020	q31724		
OCCURRENCE 18	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.31020)	q29020	q31733	q31735	q31030	q31734		
OCCURRENCE 19	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.31020)	q29020	q31743	q31745	q31040	q31744		
OCCURRENCE 20	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.31020)	q29020	q31753	q31755	q31050	q31754		
OCCURRENCE 21	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.31020)	q29020	q31763	q31765	q31060			
OCCURRENCE 22	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.31020)	q29020	q31773	q31775	q31070		q31090	F01020_
OCCURRENCE 23	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.31020)	q29020	q31783	q31785	q31080		q31090	F01020_
OCCURRENCE 24	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.31020)	q29020	q31803	q31795	q31100			OUIINON
OCCURRENCE 25	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.44020)	q43020	q44723	q44725	q44020	q44724		
OCCURRENCE 26	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.44020)	q43020	q44733	q44735	q44030	q44734		
OCCURRENCE 27	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.44020)	q43020	q44743	q44745	q44040	q44744		
OCCURRENCE 28	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.44020)	q43020	q44753	q44755	q44050	q44754		
OCCURRENCE 29	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.44020)	q43020	q44763	q44765	q44060	q44764		
OCCURRENCE 30	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.44020)	q43020	q44773	q44775	q44070		q44090	F01020_
OCCURRENCE 31	SDStrial	@SYS(SUBJID)	@SYS(SECTIONNAME.44020)	q43020	q44783	q44785	q44080		q44090	F01020_

Figure 2 : Aperçu du paramétrage de la table « VS » pour l’export sous CSExport

Le DOB mis à jour a servi de base à la création de la structure des données. L'eCRF qui avait été *designé* sur Ennov Clinical a quant à lui servi à remplir la structure en renseignant les champs correspondant aux données à exporter.

3.1.4.2. Utilisation de l'outil d'incrémentation

La différence majeure avec les tables d'export horizontales est le fait que les tables CDISC contiennent peu de variables (entre 10 et 20) mais beaucoup d'occurrences. Dès lors, le remplissage des tables d'export peut répondre à un système d'incrément afin de remplir automatiquement les occurrences avec les champs correspondant. Par exemple si dans le design de l'eCRF les champs de la page de l'examen clinique vont de 10 à 100 avec un pas de 10, il est possible de remplir la variable « VSORRES » une première fois pour la pré-inclusion et d'appliquer 9 fois un incrément de 10 pour remplir automatiquement les champs suivants.

Puis, si l'examen clinique est répété à l'identique à l'inclusion, il suffit de repérer le pas qu'il y a entre le 1^{er} champ de l'examen clinique à la pré-inclusion et celui du 1^{er} à l'inclusion (ce pas correspond à 1000 fois le nombre de pages qui séparent les deux pages dans l'eCRF) et d'incrémenter cette fois le bloc de champs de ce même nombre comme le montre la Figure 3. Ainsi nous avons les données de l'examen clinique à la pré-inclusion puis à l'inclusion, et nous pouvons continuer ainsi de suite. Cette méthode d'incrémentation est d'autant plus efficace que les pages se répètent à l'identique dans l'eCRF et avec le même nombre de pages d'écart.

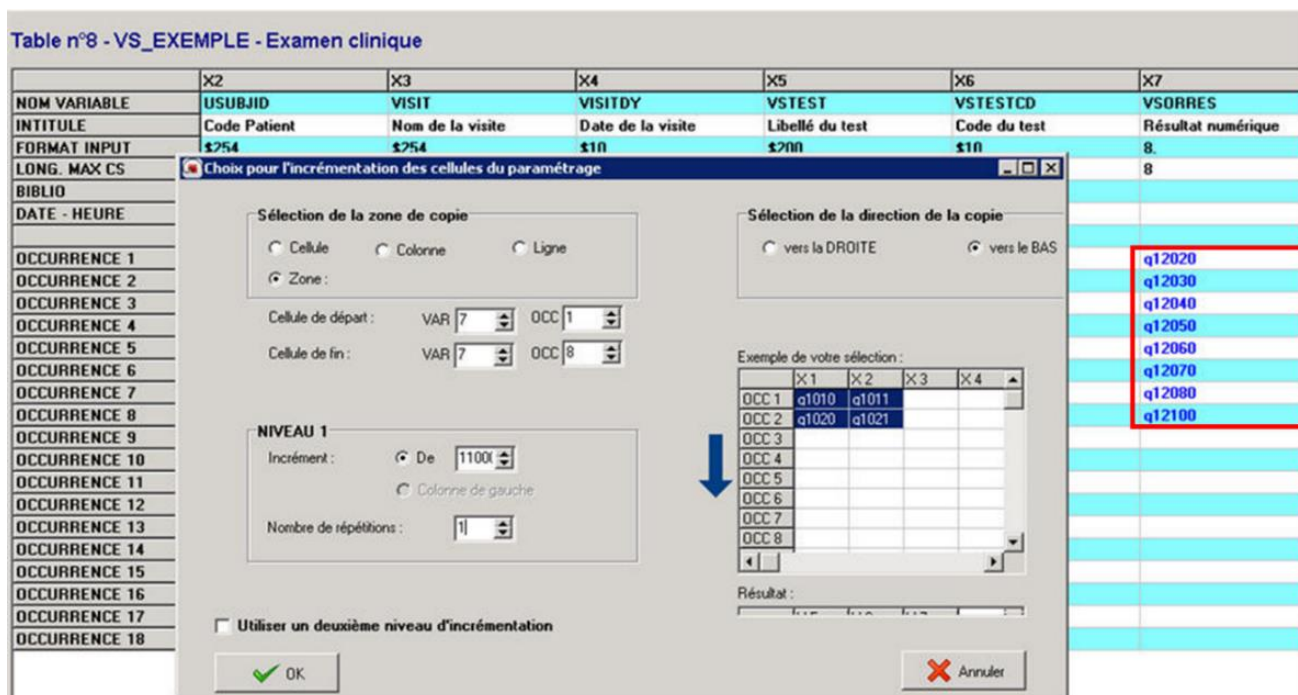


Figure 3 : Incrément d'un bloc de champs pour le paramétrage des tables d'export

3.1.4.3. Vérification des tables créées

Il existe deux moyens sous Ennov Clinical de vérifier les données renseignées dans les tables d'export. Dans un premier temps un CRF annoté peut être généré par Ennov Clinical à l'aide

du Data Handling Manual (DHM). Il permet de voir, pour chaque champ de l'eCRF, dans quelle variable (et optionnellement dans quelle table) il est exporté. Cela permet de vérifier que toutes les données nécessaires à l'analyse se trouvent bien dans les tables exportées.

Dans un second temps, lorsque des patients-tests ont été saisis, il est possible d'exporter les tables au format Excel ou SAS. Cela permet de tester l'export et de corriger d'autres erreurs éventuelles dans la définition des tables, comme par exemple une faute de frappe dans une formule qui ne donne pas le résultat attendu ou une variable mal décodée.

3.2. Programmation

Les tables CDISC ne sont pas exploitables par les programmes standards utilisés par les statisticiens de l'USMR pour leurs analyses. Un programme permettant la transposition automatique des tables dans un format plus classique a été développé, de façon à ne pas avoir à toucher aux programmes standards.

3.2.1. Définition du cahier des charges

Les besoins des statisticiens en termes de tables fournies pour l'analyse ont été définis dans un cahier des charges avec Rémi Sitta (coordonnateur métier et statisticien de l'étude SOStrial) pour interlocuteur principal. Trois priorités ont été identifiées :

- récupérer des tables avec une ligne par patient
- que les variables codées soient numériques et qu'un format leur soit associé lorsque cela est nécessaire (variables Oui/Non par exemple)
- que les libellés/noms des variables permettent d'identifier la visite à laquelle elles réfèrent

3.2.2. Logiciel utilisé

SAS (Statistical Analysis System) est un logiciel statistique et de traitement de données utilisé au sein de l'USMR par les DM et les statisticiens. Il s'agit également d'un langage de programmation à part entière. Les programmes standards SAS utilisés au sein de l'unité ont en partie fait l'objet d'une certification et le CDISC repose sur des formats gérés par SAS. Pour ces raisons et dans une démarche qualité, il s'agit donc du logiciel choisi pour réaliser le travail de programmation.

3.2.3. Etapes du développement

La programmation sous SAS a procédé par étapes de façon à complexifier petit à petit le programme jusqu'à ce qu'il soit entièrement automatisé. Les étapes sont détaillées ci-après.

3.2.3.1. Programmation sur un domaine pour une visite

Un seul domaine (« VS ») et une seule visite (la visite de pré-inclusion) ont d'abord été traités. La table est transposée pour chacune des n variables CDISC que nous souhaitons garder. Les n

tables obtenues sont ensuite fusionnées pour former la table finale. En l'occurrence, les transpositions ont été effectuées pour les variables « VSORRES » (résultat original), « VSORRESU » (unité du résultat) et « VSTEXT » (précision en rapport avec une observation). Ces transpositions ont été effectuées à l'aide d'un `proc transpose` en suivant l'exemple ci-dessous :

```
proc transpose data=VS out=transpo_VSORRES (drop= _label_
_name_) suffix=_R;
  id VSTESTCD;
  idlabel VSTEST;
  var VSORRES;
  by STUDYID USUBJID ARM ARMCD;
run;
```

Les clauses `id` et `idlabel` permettent de récupérer automatiquement les noms et labels des variables de la nouvelle table. Ils sont respectivement définis comme les valeurs des variables CDISC « VSTEST » et « VSTESTCD ». Un suffixe est ajouté aux noms des variables pour distinguer celles relevant du résultat, de l'unité ou de la précision. Les trois tables transposées sont ensuite fusionnées selon la variable `USUBJID` à l'aide de l'instruction `merge` dans une étape `data` en suivant l'exemple ci-dessous :

```
data VS_transpo_preinclusion;
  merge transpo_VSORRES
        transpo_VSORRESU
        transpo_VSTEXT ;
  by USUBJID ;
run;
```

3.2.3.2. Programmation sur un domaine pour toutes les visites

Cette procédure est généralisée pour chacune des visites apparaissant dans le domaine. Pour ce faire, une table « Liste_visites » est créée à l'aide d'un `proc sql` pour isoler l'ensemble des visites distinctes. L'instruction `call symput` appelée dans une étape `data` permet de stocker les valeurs des visites (« preinclusion », « inclusion », « suiviT3M », etc.) dans des macro-variables `&vis1`, `&vis2`, `&vis3`, etc. Le nombre de visites est également stocké dans une macro-variable `&nb_vis`. Ces macro-variables permettent d'éclater la table « VS » en `k` tables, où `k=&nb_vis` à l'aide d'une étape `data`.

Une boucle permet de faire tourner les instructions précédemment décrites pour chacune des `k` tables récupérées. Les tables transposées ainsi que les noms des variables sont préfixées avec le nom de la visite, dans un souci de traçabilité des données à l'aide du paramètre `prefix` du `proc transpose`. Cette étape permet de créer `n` fois `k` tables (où `n` est le nombre de variables CDISC à garder et `k` le nombre de visites dans le domaine). Un nouveau `merge` permet de fusionner l'ensemble des tables transposées et ce, pour l'ensemble des visites afin d'obtenir une table complète avec une ligne par individu.

3.2.3.3. Gestion des formats et des labels

La structure verticale du CDISC et le recours à un CDISC-*like* ne permettent pas de paramétrer automatiquement les formats dans l'export sous Ennov Clinical. Le choix a été fait de paramétrer l'export de façon à ce que la variable « VSORRES » soit au format numérique et de rajouter une variable « VSFMT » qui renseigne le format de la valeur de « VSORRES » le cas échéant.

Un outil de requête intégré dans le module d'export d'Ennov Clinical permet de récupérer un fichier Excel récapitulatif tous les formats utilisés dans l'eCRF et leurs modalités. Le fichier obtenu est importé dans le programme à l'aide d'un `proc import`. La requête SQL (Structured Query Language) à lancer est de la forme :

```
SELECT AUTHORIZED_VALUES, CODE, VALUE_LG2
FROM BORD0016D.ASSEMBLY
INNER JOIN BORD0016D.LIBRARIES
ON AUTHORIZED_VALUES=CS_NAME
```

Dans le programme développé, les noms des variables (« VSTESTCD »), des formats (« VSFMT ») et les libellés (« VSTEST ») sont stockés dans des macro-variables (respectivement `&obs`, `&format` et `&lib`), ainsi que leur nombre, de la même manière que les noms des visites ont été récupérés. Une étape `data` permet d'attribuer les formats récupérés pour les variables en ayant un.

De la même façon, les libellés originaux sont récupérés et le nom de la visite est ajouté devant pour définir le label. Une expression est également ajoutée pour savoir s'il s'agit du résultat, de l'unité ou d'une précision. Par exemple si le libellé original est « Poids », le libellé de l'unité à l'inclusion est : « Inclusion Poids (unité) ».

3.2.3.4. Généralisation du programme

Le paramétrage du programme permet son automatisation et sa généralisation à tous les domaines. Pour ce faire, des macro-variables sont déclarées au début du programme afin de renseigner les différents domaines que l'on souhaite transposer et, pour chacun de ces domaines, les variables CDISC que l'on souhaite garder ainsi que les suffixes qui seront ajoutés aux noms des variables et les précisions qui seront ajoutées aux labels.

Toujours dans un souci d'automatisation du programme, les tables exportées grâce au module d'export d'Ennov Clinical sont importées en faisant appel aux fichiers SAS générés par le module à l'aide de la commande `%include`.

Après la définition des paramètres et l'import des tables et des formats, une boucle permet de faire tourner les instructions décrites précédemment pour chacun des domaines dans l'ordre suivant : définition des macro-variables utilisées par le programme (`&vis`, `&var`, `&lib`, `&format`), éclatement de la table d'origine en autant de tables qu'elle contient de visites, transposition de chacune des tables obtenues pour chacune des variables paramétrées, fusion

des différentes tables et attribution des formats et des nouveaux labels. Un export des tables finalement obtenues est ensuite effectué et la session SAS est nettoyée.

3.2.4. Optimisation du programme

L'optimisation du programme repose principalement sur deux points : la modularité du programme et l'optimisation de ses performances.

3.2.4.1. Modularité du programme

Le programme final, appelé Croche-patte, est modulable. C'est-à-dire qu'un moteur effectue la transposition à l'aide du `proc transpose` et des différents `merge` comme présenté ci-avant et que des dérivées et des modules optionnels permettent d'ajouter des formats et/ou libellés, de supprimer les colonnes vides créées par le programme ou d'obtenir des tables avec une ligne par patient et par visite (au lieu d'une ligne par patient). Les codes du moteur, de l'import des tables à transposer et des différents modules et dérivées sont stockés dans des fichiers qui sont appelés à la suite du renseignement des paramètres à l'aide de la fonction `%include`. Le détail du contenu du moteur et des dérivées est explicité dans la partie « Résultats » (section 4.2.2.2, page 28).

3.2.4.2. Optimisation des performances

Le programme initialement développé contenant de nombreuses boucles imbriquées, sa structure a été revue afin d'éviter les pertes de performance. Les parties de code modifiées concernent en particulier l'application des formats et des libellés, qui testait toutes les combinaisons possibles de noms et de labels, y compris celles n'existant pas dans les données. Le code développé nécessitait également la définition de nombreuses macro-variables, en particulier pour l'application des libellés. Pour arriver au même résultat tout en étant plus cohérent avec les données réelles et en appliquant moins de boucles, l'étape `data` contenant trois boucles a été remplacée à l'aide d'une combinaison d'un `proc contents` et d'une étape `data` contenant différents `call execute` selon qu'il s'agissait de remplacer les labels, les noms ou les formats des variables. L'algorithme du programme obtenu est présenté dans la partie « Résultats » (section 4.2.2.2, page 28) et le programme complet est disponible en annexe (ANNEXE 6 : Code du programme Croche-patte, page 60).

Afin d'évaluer l'optimisation du programme, le temps d'exécution des programmes avant et après optimisation est mesuré. L'heure en début de programme est stockée dans une macro variable, de même à la fin du programme. Le temps d'exécution est ainsi défini par la différence entre ces deux macros variables et est affiché dans le rapport présenté ci-après.

3.2.5. Edition d'un rapport et notice explicative

A la fin de son exécution, si tout s'est bien déroulé, le programme édite un rapport au format PDF (*Portable Document Format*) stocké dans le dossier d'où le programme a été lancé. Le contenu de ce rapport est détaillé dans la partie « Résultats » (section 4.2.2.3, page 30). Le code

utilisé a été récupéré dans un programme utilisé à l'USMR. Il permet de rejouer la *log* pour l'imprimer dans le fichier PDF et également de faire remonter les erreurs éventuelles en début de document.

Une requête SQL permet d'éditer dans le rapport les dictionnaires de données des tables transposées à l'aide de la librairie `sashelp`. Elle est de la forme :

```
select memname, name, type, length, varnum, label, format,
informat
from sashelp.Vcolumn
where libname="TRANSP" and memname="T_&&tab&i.";
```

Une notice explicative permet d'expliquer le but et l'utilisation du programme ainsi que les contraintes nécessaires à son exécution. Cette notice est détaillée dans la partie « Résultats » (section 4.2.3, page 32).

3.3. Validation du programme

3.3.1. Authentification du programme

Le programme final fonctionne de la façon suivante : un programme principal (*main*) inclut les différents modules de code conditionnellement au paramétrage qui a été effectué. Une autre condition est ajoutée afin d'authentifier le code inclus dans le programme. Cette condition repose sur le hachage des fichiers à l'aide de l'algorithme MD5 (*Message Digest 5*).

L'empreinte numérique globale de l'ensemble des fichiers du programme Croche-patte est calculée à l'aide d'un exécutable disponible à l'USMR. Le *main* permet de recalculer l'empreinte des fichiers au moment où le programme est lancé et de la comparer avec celle de référence. Si l'empreinte diffère, le programme s'arrête et un message d'erreur est affiché dans la *log*.

3.3.2. Qualification

Le protocole de qualification est un document qualité qui décrit les différents scénarios envisagés et comment ils seront appliqués afin de vérifier la validité du programme développé. Les données sur lesquels les tests sont effectués y sont décrites. Il précise également à quoi sert le programme et quels sont ses prérequis.

L'application du protocole de qualification donne lieu à l'édition d'un rapport de qualification qui prouve le bon déroulement de chaque scénario et valide alors le programme.

4. RESULTATS

4.1. Mise en place du CDISC sur SOStrial

4.1.1. SDTM

Le SDTM construit en aveugle du travail d'Olivier Quintin pour l'étude SOStrial contient 14 domaines, définis dans le Tableau 3 ci-dessous. Le SDTM complet est disponible en annexe (ANNEXE 2: SDTM créé pour l'étude SOStrial, page 49) et le dictionnaire de données pour chacun des domaines également (ANNEXE 3 : Variables par domaine du SDTM créé pour SOStrial, page 53) . Les domaines « DM », « CM », « MH » et « AE » ont une structure horizontale tandis que les autres répondent au principe du question/réponse explicité dans le contexte à l'aide des variable « --TEST » et « -ORRES ».

Tableau 3 : Domaines du SDTM créés pour l'étude SOStrial en aveugle du travail déjà effectué

Domaine	Description
Adverse Event (AE)	Informations concernant les événements indésirables survenus au cours de l'étude.
Concomitant/prior Medications (CM)	Traitements pris par le participant, que ce soit des traitements pré-spécifiés dans le CRF (par exemple : « Le patient prend-il de la Mémantine ? ») ou non (page de traitements concomitants en fin de CRF).
Death Details (DD)	Détails sur le décès du participant.
Demographics (DM)	Informations sociodémographiques du participant (date de naissance, sexe) et des informations en rapport avec l'étude comme le statut de randomisation, la date d'inclusion ou la date de fin d'étude
Drug Accountability (DA)	Informations liées à la compliance du traitement de l'étude par le participant tout au long des visites.
Exposure (EX)	Informations relatives à l'administration du traitement
Medical History (MH)	Antécédents médicaux du participant, qu'ils soient pré-spécifiés dans le CRF ou non.
Physical Examination (PE)	Informations ayant trait à des examens physiques effectués lors des visites.
Questionnaires (QS)	Informations concernant les auto-questionnaires ou les échelles de mesure effectués lors des visites.
Subject Characteristics (SC)	Caractéristiques du participant (principalement des données sociodémographiques non prévues dans DM).

Domaine	Description
Subject Status (SS)	Informations concernant le statut du participant et répétées à chaque visite (par exemple « Le participant est-il en institution ? »).
Subject Visits (SV)	Informations ayant trait aux visites effectuées par le participant (dates, raison pour laquelle elle n'a pas été faite, accompagnant, etc.).
Trial Inclusion / exclusion criteria (TI)	Critères d'inclusion, de non-inclusion et d'exclusion de l'étude.
Vital Signs (VS)	Informations ayant trait à des signes vitaux recueillis lors des visites.

4.1.2. DOB

Le DOB créé par Olivier Quintin et ajusté comme expliqué dans la méthode contient 9 domaines. Le Tableau 4 ci-après résume le contenu de chacun. Le détail des variables des domaines peut être retrouvé dans l'impression du paramétrage des tables (*cf.* section 4.1.3, page 25).

Tableau 4 : Domaines créés pour le DOB de l'étude SOStrial une fois la mise en commun effectuée

Domaine	Description
Adverse Event (AE)	Informations concernant les événements indésirables survenus au cours de l'étude.
Concomitant/prior Medications (CM)	Traitements pris par le participant, que ce soit les psychotropes ou les traitements non spécifiés.
Disposition (DS)	Informations liées à l'étude (consentement, randomisation, fin d'étude, etc.)
Exposure (EX)	Informations relatives à l'administration des IAC
Inclusion / exclusion criteria (IE)	Critères d'inclusion, de non-inclusion et d'exclusion de l'étude.
Medical History (MH)	Antécédents médicaux du participant.
Questionnaires (QS)	Informations concernant le MMSE et l'échelle de Katz.
Subject Characteristics (SC)	Données sociodémographiques, caractéristiques et statut du participant.
Vital Signs (VS)	Informations ayant trait à des signes vitaux recueillis lors des visites.

4.1.3. Paramétrage des tables d'export

Les tables créées pour l'export respectent le DOB présenté ci-avant. Elles sont donc au nombre de 9 et contiennent entre 11 et 19 variables et entre 7 et 216 occurrences. Le module CSEXPOT d'Ennov Clinical permet d'imprimer la liste des tables et leur paramétrage. La Figure 4 présente la liste des tables. Le détail du paramétrage de toutes les tables est fourni en annexe (ANNEXE 4 : Paramétrage des tables sous CSEXPOT, page 56).

Liste des tables d'export

BORD0016 - CHUBX 2016/27

---- Informations ----

Etude : BORD0016
Libname : CDISC
Chemin : R:\EQUIPES\USMR\DATA MANAGEMENT\Stagiaire\Pauline GAFFEZ\CDISC_SOS_TRIAL\Programmation\Croche-patte - v0.1 - 31052019\tables

#	Table	Description	Fichier	Dat	Nb var.	Nb occ.	Verr.	DHM	Par
1	VS	Examen clinique	VS	VS	17	56	NON	OUI	PAULINE GAFFEZ (3130-196853)
2	QS	Questionnaires	QS	QS	18	216	NON	OUI	PAULINE GAFFEZ (3130-196853)
3	SC	Caractéristiques du sujet	SC	SC	17	43	NON	OUI	PAULINE GAFFEZ (3130-196853)
4	IE	Critères d'inclusion/exclusion	IE	IE	16	22	NON	OUI	PAULINE GAFFEZ (3130-196853)
5	CM	Traitements concomitants	CM	CM	19	92	NON	OUI	PAULINE GAFFEZ (3130-196853)
6	DS	Suivi du sujet	DS	DS	17	39	NON	OUI	PAULINE GAFFEZ (3130-196853)
7	AE	Effets indésirables	AE	AE	17	40	NON	OUI	PAULINE GAFFEZ (3130-196853)
9	MH	Antécédents	MH	MH	11	25	NON	OUI	PAULINE GAFFEZ (3130-196853)
10	EX	Traitement à l'étude	EX	EX	15	7	NON	OUI	PAULINE GAFFEZ (3130-196853)

Figure 4 : Impression de la liste des tables paramétrées dans le module d'Ennov Clinical CSEXPOT

4.1.4. Tables d'export

CSEXPOT permet d'exporter les tables au format Excel ou SAS (entre autres). Dans le deuxième cas, un fichier .dat contenant les données et un fichier .sas contenant le paramétrage des variables sont créés. Le second fichier crée une table et appelle le premier pour remplir la table créée. Un troisième exécutable (.bat) permet de charger la table directement sur SAS. La Figure 5 ci-après représente un extrait du résultat obtenu une fois la table ouverte sous SAS.

	Nom de l'étude	Domaine	Code Patient	Número du centre	Nom de la visite	Número de la visite	Número de séquence	Date de la visite	Date de la visite	Groupe de randomisation	Code de randomisation	Libellé du test	Code du test	Résultat numérique	Résultat caractères	Unité	Précision	Format de VSORRES
1849	SOStrial	VS	020033	2	PRE-INCLUSION	1	1	04/03/2019	04/03/2019	1 - Poursuite des IAC	23369	Poids	PDS	50		KG		
1850	SOStrial	VS	020033	2	PRE-INCLUSION	1	2	04/03/2019	04/03/2019	1 - Poursuite des IAC	23369	Pression artérielle Systolique mmHg	PAS	150		MMHG		
1851	SOStrial	VS	020033	2	PRE-INCLUSION	1	3	04/03/2019	04/03/2019	1 - Poursuite des IAC	23369	Pression artérielle Diastolique mmHg	PAD	100		MMHG		
1852	SOStrial	VS	020033	2	PRE-INCLUSION	1	4	04/03/2019	04/03/2019	1 - Poursuite des IAC	23369	Fréquence cardiaque /min	FC	80		/MIN		
1853	SOStrial	VS	020033	2	PRE-INCLUSION	1	5	04/03/2019	04/03/2019	1 - Poursuite des IAC	23369	Taille cm	TAILLE	160		CM		
1854	SOStrial	VS	020033	2	PRE-INCLUSION	1	6	04/03/2019	04/03/2019	1 - Poursuite des IAC	23369	Résultat examen clinique	EC	1	Normal			F01020_
1855	SOStrial	VS	020033	2	PRE-INCLUSION	1	7	04/03/2019	04/03/2019	1 - Poursuite des IAC	23369	Résultat examen neurologique	EN	1	Normal			F01020_
1856	SOStrial	VS	020033	2	PRE-INCLUSION	1	8	04/03/2019	04/03/2019	1 - Poursuite des IAC	23369	Le test urinaire de grossesse est-il négatif	TDG	1	Oui			QUINON
1857	SOStrial	VS	020033	2	INCLUSION	2	9	06/03/2019	06/03/2019	1 - Poursuite des IAC	23369	Poids	PDS	60		KG		
1858	SOStrial	VS	020033	2	INCLUSION	2	10	06/03/2019	06/03/2019	1 - Poursuite des IAC	23369	Pression artérielle Systolique mmHg	PAS	150		MMHG		
1859	SOStrial	VS	020033	2	INCLUSION	2	11	06/03/2019	06/03/2019	1 - Poursuite des IAC	23369	Pression artérielle Diastolique mmHg	PAD	100		MMHG		
1860	SOStrial	VS	020033	2	INCLUSION	2	12	06/03/2019	06/03/2019	1 - Poursuite des IAC	23369	Fréquence cardiaque /min	FC	100		/MIN		
1861	SOStrial	VS	020033	2	INCLUSION	2	13	06/03/2019	06/03/2019	1 - Poursuite des IAC	23369	Taille cm	TAILLE			CM		
1862	SOStrial	VS	020033	2	INCLUSION	2	14	06/03/2019	06/03/2019	1 - Poursuite des IAC	23369	Résultat examen clinique	EC	1	Normal		,HFGHK	F01020_
1863	SOStrial	VS	020033	2	INCLUSION	2	15	06/03/2019	06/03/2019	1 - Poursuite des IAC	23369	Résultat examen neurologique	EN	0	Anormal		,HFGHK	F01020_
1864	SOStrial	VS	020033	2	INCLUSION	2	16	06/03/2019	06/03/2019	1 - Poursuite des IAC	23369	Le test urinaire de grossesse est-il négatif	TDG	1	Oui			QUINON
1865	SOStrial	VS	020033	2	SUIVI T6M RANDOMISATIO	4	17	08/03/2019	08/03/2019	1 - Poursuite des IAC	23369	Poids	PDS	54		KG		
1866	SOStrial	VS	020033	2	SUIVI T6M RANDOMISATIO	4	18	08/03/2019	08/03/2019	1 - Poursuite des IAC	23369	Pression artérielle Systolique mmHg	PAS	150		MMHG		
1867	SOStrial	VS	020033	2	SUIVI T6M RANDOMISATIO	4	19	08/03/2019	08/03/2019	1 - Poursuite des IAC	23369	Pression artérielle Diastolique mmHg	PAD	100		MMHG		
1868	SOStrial	VS	020033	2	SUIVI T6M RANDOMISATIO	4	20	08/03/2019	08/03/2019	1 - Poursuite des IAC	23369	Fréquence cardiaque /min	FC	100		/MIN		
1869	SOStrial	VS	020033	2	SUIVI T6M RANDOMISATIO	4	21	08/03/2019	08/03/2019	1 - Poursuite des IAC	23369	Taille cm	TAILLE			CM		
1870	SOStrial	VS	020033	2	SUIVI T6M RANDOMISATIO	4	22	08/03/2019	08/03/2019	1 - Poursuite des IAC	23369	Résultat examen clinique	EC	0	Anormal		THGJF	F01020_
1871	SOStrial	VS	020033	2	SUIVI T6M RANDOMISATIO	4	23	08/03/2019	08/03/2019	1 - Poursuite des IAC	23369	Résultat examen neurologique	EN	1	Normal		THGJF	F01020_
1872	SOStrial	VS	020033	2	SUIVI T6M RANDOMISATIO	4	24	08/03/2019	08/03/2019	1 - Poursuite des IAC	23369	Le test urinaire de grossesse est-il négatif	TDG	K	K			QUINON

Figure 5 : Extrait du résultat de la table VS de l'étude SOStrial importée dans SAS

4.2. Programmation

4.2.1. Cahier des charges

Le cahier des charges complet dressé pour les tables à fournir aux statisticiens de l'USMR dans le cadre d'une étude réalisée en CDISC a été mis en annexe (ANNEXE 5 : Cahier des charges pour des tables fournies pour l'analyse, page 59). Ce cahier des charges est en attente de validation conformément au SMQ. Les contraintes sont les suivantes :

- La programmation doit être effectuée sous SAS.
- Le programme doit être étude-indépendant, il doit fonctionner quelles que soient les tables exportées.
- Le paramétrage du programme doit être relativement simple.
- Il ne faut pas que le paramétrage de l'eCRF et de l'export devienne trop complexe.

Les besoins des statisticiens sont définis comme ceci :

- La priorité est donnée pour avoir toutes les données en lignes dans les tables (une ligne par individu). Il faut donc transposer les tables CDISC tout en gardant l'information de la visite d'où provient la donnée, par exemple grâce à l'ajout d'un préfixe. Ensuite, si cela est réalisable, les statisticiens peuvent avoir besoin des données pour toutes les visites les unes à la suite des autres dans une même table (une ligne par visite et par individu).
- Il faut récupérer les observations associées à des bibliothèques au format numérique avec leur format pour pouvoir décoder les variables sous SAS. Des essais seront menés pour voir quelle solution est la plus efficiente afin de récupérer les formats des variables avec les tables au format CDISC. Eventuellement, des formats peuvent être créés même pour les variables numériques, car en pratique les statisticiens peuvent être amenés à le faire.
- Les labels des variables doivent être gardés car ce sont eux qui apparaissent dans les rapports d'analyse. Les noms donnés aux variables sont exploités dans les programmes et ne sont pas affichés.
- Il apparaît nécessaire de définir des conventions de nommage, tant pour les noms des tables, que des variables, que des noms des visites. A cette fin, la convention de nommage existante pourra être mise à jour et complétée afin de standardiser les pratiques.

4.2.2. Programme(s)

4.2.2.1. Développement

Pour illustrer la progression de la programmation entre les différentes phases, seul le passage concernant le `proc transpose` est détaillé ici.

En premier lieu, le code présenté dans la section 3.2.3.1, page 18 est utilisé sur la table « VS » pour la visite de pré-inclusion uniquement. Cette transposition s'effectue pour les trois variables « VSORRES », « VSORRESU » et « VSTEXT » séparément.

Dans un second temps les variables CDISC sont déclarées en paramètres dans des macros-variables et une boucle est mise en place pour automatiser la transposition sur chaque variable :

```
%do j=1 %to &nbvar ;
  proc transpose data=preinclusion
    out=%%var&j.._preinclusion (drop=_label_ _name_)
    prefix=preinclusion_;
    id VSTESTCD;
    idlabel VSTEST;
    var %%var&j;
    by STUDYID USUBJID VISITDY ARM ARMCD;
  run;
%end ;
```


Ensuite les visites sont récupérées dans des macro variables `&&vis&l` comme expliqué dans la méthode, ce qui permet de faire tourner le code sur toutes les visites :

```
%do j=1 %to &nbvar.;
  %do l=1 %to &nbvis.;
    proc transpose data=&&vis&l. out=&&var&j.._&&vis&l.
      (drop=_label_ _name_) prefix=&&vis&l.._;
      id VSTESTCD;
      idlabel VSTEST;
      var &&var&j.;
      by STUDYID USUBJID ARM ARMCD;
    run;
  %end ;
%end ;
```

Finalement, les noms des tables sont paramétrés dans des macros-variables pour pouvoir faire tourner le programme obtenu sur l'ensemble des tables que l'on souhaite transposer :

```
%do i=1 %to &nbtabs. ;
  %do j=1 %to &nbvar.;
    %do l=1 %to &nbvis. ;
      proc transpose data=&&tab&i.._&&vis&l.
        out=&&var&j.._&&vis&l. (drop=_label_ _name_)
        prefix=&&vis&l.;
        id &&tab&i..TESTCD;
        idlabel &&tab&i..TEST;
        var &&var&j.;
        by STUDYID USUBJID ARM ARMCD;
      run ;
    %end ;
  %end ;
%end ;
```

4.2.2.2. Optimisation

Lors de l'étape d'optimisation le programme Croche-patte a été modulé puis les boucles ont été remplacées comme expliqué dans la méthode. Les dérivées finalement développées sont les suivantes :

- 1) Export brut des tables transposées sans autre ajout
- 2) Application des formats puis export des tables transposées
- 3) Application des libellés puis export des tables transposées
- 4) Application des formats et des libellés puis export des tables transposées

Deux modules sont également disponibles :

- A) Avant application de la dérivée : transformation des tables pour avoir une ligne par patient et par visite (contre une ligne par patient sinon)

B) Après application de la dérivée : suppression de toutes les colonnes vides (celles créées par le programme mais potentiellement aussi celles ne contenant aucune donnée)

Ci-après les Tableau 5 et Tableau 6 reprennent les grandes étapes concernant respectivement le moteur et les dérivées.

Tableau 5 : Algorithme rédigé du moteur du programme

Etape	Action	SAS
1	Import des tables originales	Proc import
2	Liste des visites distinctes de la table XX et nombre	Proc SQL
3	Noms des visites stockées dans les macros XX_vis1 à XX_visX	Call symput
4	Eclatement de la table en X tables selon les visites	Data output
5	Transposition de chaque table pour chacune de K variables paramétrées	Proc transpose
6	Regroupement de toutes les tables	Data merge

Tableau 6 : Algorithme rédigé des dérivées du programme

Dérivée	Etape	Action	SAS
1 : export brut	1	Création d'une bibliothèque	Libname
	2	Copie des tables de la work vers la librairie créée	Proc copy
	3	Nettoyage de la work	Proc datasets
2 : application des formats	1	Import du fichier format et définition des formats SAS	Proc import Proc format
	2	Liste des variables avec formats et des formats	Proc SQL
	3	Noms (resp. formats) des variables stockés dans des macros XX_obs1 à XX_obsX (resp. XX_fmt1 à XX_fmtX)	Call symput
	4	Suppression de tous les formats	Data format

Dérivée	Etape	Action	SAS
	5	Application des formats aux variables concernées	Proc contents Data call execute
	6	Etapes 1,2,3 de la dérivée 1	Cf. dérivée 1
3 : application des libellés	1	Application des libellés	Proc contents Data call execute
	2	Etapes 1,2,3 de la dérivée 1	Cf. dérivée 1
4 : application des formats et des libellés	1	Etapes 1 à 5 de la dérivée 2 puis étapes 1 et 2 de la dérivée 3	Cf. dérivées 2 et 3

Les temps d'exécution avant et après optimisation du programme sont donnés par le Tableau 7. Ces résultats sont commentés dans la discussion, section 5.2.1 page 41.

Tableau 7 : Temps d'exécution des dérivées 2, 3 et 4 avant et après optimisation

Dérivée	Temps d'exécution (en secondes)	
	Avant optimisation	Après optimisation
2 : application des formats	2,60	4,13
3 : application des libellés	2,72	2,62
4 : application des formats et des libellés	3,03	4,50

4.2.2.3. Rapport PDF

Le rapport édité par le programme Croche-patte contient :

- Des informations concernant la version du programme et la dérivée utilisée
- Certains des paramètres utilisés (dossier, tables traitées, variables CDISC traitées)
- Le temps d'exécution du programme
- Les erreurs apparues dans la *log* le cas échéant
- Les dictionnaires de données des tables obtenues
- Une impression de la *log* (avec une impression de toutes les macro-variables définies par le programme)

Un exemple de rapport édité par le programme est fourni dans la Figure 6 ci-après (la partie avec l'impression de la *log* a volontairement été tronquée).

Programme Croche-patte v0.1 31052019

1

Rapport dérivée n°1

Export brut des tables transposées

Date et heure d'exécution: 7 juin 2019 15:26
 Dossier: R:\EQUIPES\USMR\DATA MANAGEMENT\Stagiaire\Pauline GAFFEZ\CDISC_SOS_TRIAL\Programmation\Croche-patte - v0.1 - 31052019\tables
 Bibliothèque d'export: transpo
 Durée d'exécution: inférieure à 1 secondes
 Tables traitées: 'VS'
 Variables communes à toutes les tables: STUDYID USUBJID ARM ARMCD VISIT VISITNUM VISITDY
 Les variables CDISC traitées sont les suivantes:
 VSORRES, VSTEXT, VGORRESU
 Module A utilisé: 1 (0=non 1=oui)
 Module B utilisé: 1 (0=non 1=oui)

Dictionnaire de données de la table T_VS

2

Member Name	Column Name	Column Type	Column Length	Column Number in Table	Column Label	Column Format	Column Informat
T_VS	STUDYID	char	75	1	Nom de l'étude	\$75.	
T_VS	USUBJID	char	254	2	Code Patient	\$254.	
T_VS	ARM	char	254	3	Groupe de randomisation	\$254.	
T_VS	ARMCD	char	254	4	Code de randomisation	\$254.	
T_VS	VISIT	char	254	5	Nom de la visite	\$254.	
T_VS	VISITNUM	num	8	6	Numéro de la visite	2.	
T_VS	VISITDY	char	10	7	Date de la visite	\$10.	
T_VS	PDS_R	num	8	8	Poids	8.	
T_VS	PAS_R	num	8	9	Pression artérielle Systolique mmHg	8.	
T_VS	PAD_R	num	8	10	Pression artérielle Diastolique mmHg	8.	
T_VS	FC_R	num	8	11	Fréquence cardiaque /min	8.	
T_VS	TAILLE_R	num	8	12	Taille cm	8.	
T_VS	EC_R	num	8	13	Résultat examen clinique	8.	
T_VS	EN_R	num	8	14	Résultat examen neurologique	8.	
T_VS	TDG_R	num	8	15	Le test urinaire de grossesse est-il négatif	8.	
T_VS	EC_P	char	200	16	Résultat examen clinique	\$200.	
T_VS	EN_P	char	200	17	Résultat examen neurologique	\$200.	
T_VS	PDS_U	char	10	18	Poids	\$10.	
T_VS	PAS_U	char	10	19	Pression artérielle Systolique mmHg	\$10.	
T_VS	PAD_U	char	10	20	Pression artérielle Diastolique mmHg	\$10.	
T_VS	FC_U	char	10	21	Fréquence cardiaque /min	\$10.	
T_VS	TAILLE_U	char	10	22	Taille cm	\$10.	

Figure 6 : Exemple de rapport édité par le programme Croche-patte

4.2.2.1. Programme final

Le code des différentes parties du programme final sont disponibles en annexe (ANNEXE 6 : Code du programme Croche-patte, page 60). Pour les dérivées, seul le code de la dérivée 4 est disponible, les autres reprenant des parties de ce code. La Figure 7 ci-après schématise le fonctionnement du programme : le fichier principal (« Croche-patte ») permet de renseigner les paramètres nécessaires au fonctionnement du programme (noms des tables, noms des variables CDISC, dérivée à appliquer, etc.) et appelle les fichiers « Import », « Main » et « Rapport ». Le premier permet d'importer les tables à transposer, le second inclut les bouts de code du moteur, de la dérivée souhaitée et des modules optionnellement au paramétrage (mais aussi au résultat de l'authentification des programmes comme expliqué dans la méthode) et le dernier permet d'éditer le rapport PDF.

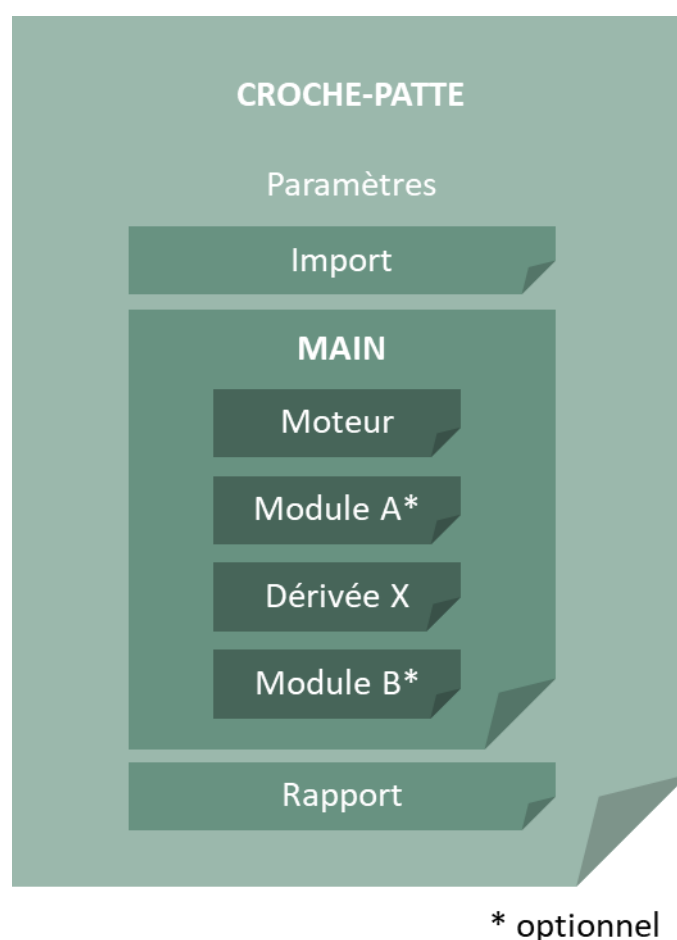


Figure 7 : Schéma du fonctionnement du programme développé Croche-patte

4.2.3. Notice explicative

La notice d'utilisation du programme est disponible en annexe (ANNEXE 7 : Notice d'utilisation du programme Croche-patte, page 71). Cette notice est en attente de validation conformément au SMQ. Elle détaille le but du programme, les conditions liées à son utilisation puis explicite les étapes à mettre en œuvre afin d'assurer le bon déroulement du programme :

- 1) La récupération du programme
- 2) Le remplissage des paramètres
- 3) Le lancement du programme

La définition des paramètres et les dérivées auxquelles ils s’appliquent sont détaillés dans le Tableau 8 ci-après.

Tableau 8 : Paramétrage du programme selon les dérivées

Noms des macros	Dérivées	Description
Tab1-Tabn	Toutes	Noms des tables que l’on souhaite transposer. Le nom de ces tables est par la suite désigné par XX (ex : tab1=VS).
Var_com	Toutes	Variabes communes à toutes les tables que l’on souhaite garder une fois les tables transposées, séparées par des espaces (ex : « STUDYID USUBJID ARM ARMCD »). Si le module A est utilisé, les variables contenant les nom et numéro de la visite doivent y figurer.
XX_var1-XX_vark	Toutes	Noms des variables CDISC que l’on souhaite transposer (ex : VS_var1=VSORRESU). Ces paramètres sont à renseigner pour chaque table à traiter.
XX_s1-XX_sk	Toutes	Suffixes ajoutés aux noms des variables (noms définis par la variable « XXTESTCD ») lors de la transposition, commençant par un « _ ». Ces paramètres doivent obligatoirement être renseignés pour chaque variable CDISC définie précédemment (ex : VS_s1=_U).
XX_lab1-XX_labk	3 et 4	Précisions ajoutées aux labels des variables (labels définis par la variable « XXTEST »). Ces paramètres doivent obligatoirement être renseignés pour chaque variable CDISC définie précédemment (ex : VS_lab1=unité).
Formats	2 et 4	Nom du fichier contenant les formats (ex : RES_ORACLE_EXCEL_20190307_135941.xls).
Module_A (ou B)	Toutes	Inclusion ou non des modules optionnels (1=oui 0=non)
Derivee	Toutes	Numéro de la dérivée que l’on souhaite lancer (1=export brut 2=formats 3=libellés 4=formats et libellés)

Les contraintes des différentes dérivées sont également explicitées dans la notice et sont résumées dans le Tableau 9 ci-après.

Tableau 9 : Contraintes à respecter pour l'utilisation du programme selon la dérivée souhaitée

Dérivées	Contraintes
Toutes	<ul style="list-style-type: none"> • Utiliser les fichiers SAS exportés par csexport (le chemin et le nom de la bibliothèque pour l'import des tables doivent être modifiés avant l'export). • Avoir une variable VISIT contenant le nom de la visite. • Avoir une variable VISITNUM contenant le numéro de la visite. • Avoir une variable USUBJID contenant l'identifiant du participant. • Avoir une variable « --TESTCD » contenant les noms des futures variables. • Avoir une variable « --TEST » contenant les libellés des futures variables.
2 et 4	<ul style="list-style-type: none"> • Avoir un fichier contenant tous les formats utilisés dans l'étude (obtenu à l'aide d'une requête dans csexport). • Avoir une variables « --FMT » contenant le format de la valeur de « --ORRES » le cas échéant. • Que la variable « --ORRES » soit préférentiellement au <u>format numérique</u> et ne contienne que des données numériques (pas de dates par exemple). Si cette contrainte n'est pas respectée, le programme fonctionne tout de même mais les formats sont appliqués au format caractère et les données restent donc au format caractère.

4.3. Validation du programme

Le protocole complet de qualification du programme a été mis en annexe (ANNEXE 8 : Protocole de qualification du programme Croche-patte, page 72). Ce protocole est en cours de validation conformément au SMQ. Les scénarios déployés pour la validation du programme sont au nombre de six et sont détaillés ci-après :

1) *Transposition et application correctes des formats et libellés*

- Exécution du fichier « Croche-patte_scenario1_v0.1_31052019.sas »
- Edition du rapport PDF sans erreur notifiée

2) *Authentification du programme grâce aux empreintes MD5*

- Exécution du script « SC02 - Ajout d'un espace.vbs »
- Exécution du fichier « Croche-patte_scenario2_v0.1_31052019.sas »
- Message d'erreur dans la log et arrêt du programme
- Exécution du script « SC02 - Suppression d'un espace.vbs »

3) *Mauvais paramétrage des tables (tab1=TS au lieu de VS)*

- Exécution du fichier « Croche-patte_scenario3_v0.1_31052019.sas »
- Edition du rapport PDF avec notification des erreurs Edition du rapport PDF avec remontée des erreurs

4) *Mauvais paramétrage du programme (derivee=5)*

- Exécution du fichier « Croche-patte_scenario4_v0.1_31052019.sas »
- Message d'erreur dans la log et arrêt du programme

5) *Jeu de données incorrect (pas de variable VISIT)*

- Exécution du fichier « Croche-patte_scenario5_v0.1_31052019.sas »
- Edition du rapport PDF avec notification des erreurs

6) *Variable --ORRES au format caractère*

- Exécution du fichier « Croche-patte_scenario6_v0.1_31052019.sas »
- Edition du rapport PDF avec notification des erreurs

Le rapport de qualification, résultat de l'application de ce protocole est également disponible en annexe (ANNEXE 9 : Rapport de qualification du programme Croche-patte, page 73). Il s'agit d'une version qui n'a pas encore été validée conformément au SMQ.

4.4. Intervention forum méthodologique

Dans le cadre de l'alternance Rodolphe Thiébaud, directeur de l'USMR, a souhaité qu'une présentation en forum méthodologique soit effectuée. Les forums méthodologiques sont des réunions ayant lieu tous les mois, auxquelles toutes les équipes de recherche de l'ISPED sont conviées et durant lesquelles différentes présentations ont lieu sur des sujets de recherche ou des thèmes actuels en recherche clinique.

Le 12 juin 2019 une présentation pour expliquer le contexte, les objectifs et l'avancement de ce stage a été effectuée. Le script de cette présentation se trouve en annexe (ANNEXE 10 : Script de l'intervention en forum méthodologique du 12/06/2019, page 75). Les échanges engendrés par cette présentation ont permis d'enrichir la discussion présentée ci-après.

5. DISCUSSION

5.1. Faisabilité et intérêt de la mise en place du CDISC au sein de l'USMR

La mise en place du CDISC sur l'étude SOStrial a permis de mettre en lumière certains points vis-à-vis de la faisabilité de sa généralisation pour la gestion des données à l'USMR. Ces points sont explicités ci-après.

5.1.1. Mise en place du CDISC sur SOStrial

5.1.1.1. Mise en place du SDTM

La mise en place du SDTM pour l'étude SOStrial a fait apparaître plusieurs points de réflexion. Le premier a été la multiplication des domaines. Au total, 13 domaines ont été définis, ce qui aurait abouti à la création de 13 tables pour l'export. Or ces domaines contiennent parfois très peu d'information, comme par exemple le domaine « Death Details (DD) » qui contient deux informations recueillies dans la fiche de fin d'étude (la cause initiale du décès et les autres causes ayant contribué au décès). Il aurait également été possible de mettre ces détails dans le domaine « Death Status », qui prévoit une variable « DSREAS » pour la raison du décès mais cela aurait aussi abouti à la création d'une nouvelle table.

Le deuxième point est que certaines observations du cahier ont eu du mal à trouver leur place au sein du SDTM. Par exemple l'information concernant l'accompagnement ou non du participant lors de la visite de suivi a mis en lumière un certain manque de souplesse du standard. Il aurait peut-être été possible d'utiliser un domaine « SUPPQUAL » pour récolter cette information supplémentaire, mais la complexité de la gestion de ce type de données n'a pas permis de le faire.

Cela amène au troisième point qui est la complexité et la difficulté d'appropriation du standard sans une formation complète. Par exemple les observations de type « Autre, préciser » peuvent trouver leur place soit dans la variable « -ORRES » directement, soit dans un domaine supplémentaire « SUPP- » qui permet de raccrocher des précisions à des observations d'un domaine. Le choix se fait en fonction de la classe de l'observation à laquelle se rattache la précision. Il en résulte que les règles d'application du standard n'ont pas été complètement respectées et que le choix a été fait d'appliquer un standard s'inspirant du CDISC pour l'étude SOStrial.

5.1.1.2. Recours à une standardisation CDISC-like

Pour ces raisons, le pur CDISC n'a pas été mis en place pour l'étude SOStrial. Ainsi, c'est une standardisation s'inspirant du CDISC qui a été choisie. Les tables horizontales ont été gardées telles quelles mais les tables verticales ont subi quelques modifications comme explicité dans la méthode. Ces modifications apportent de la souplesse au CDISC tout en gardant le principe

de verticalité à l'aide des variables « --TEST » et « --ORRES » mais impliquent certaines limites.

La première limite identifiée est la gestion des formats. Le fait de ne pas utiliser le fichier XML DEFINE qui décrit les métadonnées des tables ne permet pas de récupérer automatiquement les formats des données, en particulier lorsqu'un thesaurus est appliqué. De ce fait, l'ajout d'une variable « --FMT » est nécessaire. Cette dernière permet de préciser le thesaurus (appelé bibliothèque) du résultat contenu dans la variable « --ORRES » le cas échéant.

Toujours dans un souci de format, la variable « --ORRES », qui contient le résultat principal de l'observation, est définie au format numérique dans les tables d'export. Dans le CDISC-like qui a été appliqué, cette variable contient en effet principalement des données numériques et des codes associés à des bibliothèques. Le problème est que dans un premier temps cette variable contenait parfois également des dates et que lorsqu'une date est exportée au format numérique elle est tronquée au premier « / ». Or, pour respecter le cahier des charges établi avec les statisticiens il n'est pas possible de définir cette variable au format texte, car après transposition les données doivent être au format numérique. Le problème pourrait également se poser si des champs texte se retrouvaient dans la variable « --ORRES », même si en pratique ces champs doivent se trouver dans la variable qui a été ajoutée « --TEXT » ou dans une autre variable CDISC.

Deux solutions peuvent être proposées pour pallier au problème des dates : ajouter une variable aux domaines pour récolter les dates, comme en CDISC pur, ou bien ajouter une table contenant toutes les dates récoltées dans l'étude. Cependant, pour pouvoir gérer des dates de format différents (date complète et date partielle comme la date de naissance) il faudrait ajouter une variable « --FMTDAT » pour pouvoir rester dans un format de table verticale. Sinon, il est possible de construire la table des dates dans un format horizontal, sur le même modèle que les domaines « Adverse event » ou « Medical History » par exemple.

Ce type de limite implique la nécessité de définir des conventions avec les différents groupes métier (gestion des données et statistiques principalement) pour pouvoir définir les domaines-types à créer avec les variables qu'ils contiendraient et les formats qui seraient appliqués. Cela permettrait d'assurer la standardisation des pratiques au sein de l'équipe.

5.1.2. Implications de la mise en place du CDISC à l'USMR

5.1.2.1. Changement des pratiques et documents actuels

La mise en place d'une standardisation CDISC-like sur l'étude SOStrial a permis d'amener des réflexions plus larges sur les pratiques au sein de l'USMR. En effet, la mise en place d'un standard nécessite l'adaptation des pratiques, avec par exemple une autre gestion des formats comme expliqué plus haut. De la même manière, il est possible de remettre en question les documents actuellement mis en place dans l'unité pour la gestion des données. Par exemple, le DOB, qui est un document définissant l'organisation de la base de données, n'est peut-être plus nécessaire si les pratiques sont entièrement standardisées. En effet, une observation aurait

toujours le même nom (ou une même construction de nom) et se trouverait toujours dans la même table. Dans le cas où le DOB serait maintenu, il faudrait quoi qu’il en soit le changer pour l’adapter au SDTM, qui est un format différent du format actuel.

Il en va de même pour le CO annoté. Il s’agit d’un document reprenant le CO développé avec l’équipe investigatrice, avec les noms de variables annotés à côté de l’observation en question. Il permet de fournir aux statisticiens les noms des variables dans la base de données correspondant aux observations remplies sur le CRF. Si une convention de nommage complète est établie, alors sa pertinence peut être remise en cause. En revanche s’il est maintenu, l’outil d’Ennov Clinical (DHM) qui permet de le sortir à l’heure actuelle ne le permettrait plus sous le format CDISC. En effet, seule l’indication de la variable CDISC apparaît, et non pas le nom de l’observation. La Figure 8 illustre ce point. Il est alors impossible de distinguer deux observations et de savoir exactement où se trouve la valeur recherchée. Ce document servant aux statisticiens pour repérer les variables à analyser dans les tables, il est possible en revanche d’imprimer un dictionnaire de données à l’aide de SAS une fois que les tables ont été transformées par le programme pour le repérage des variables. C’est ce qui a été fait dans le rapport PDF édité par le programme Croche-patte.

Page 8 - CRF 1 - EXAMEN CLINIQUE

CHU
Hospitals de
Bordeaux

USMR du CHU de Bordeaux
Etude <SOS TRIAL>

Centre Participant Code lettre PRE_INC
Version n°1.0 du 06/11/2017

EXAMEN CLINIQUE

Poids VSTEST VSORRES kg VSORRESU VSTESTCD

Pression artérielle VSTEST Systolique VSORRES mm Hg VSORRESU VSTESTCD

VSTEST Diastolique VSORRES mm Hg VSORRESU VSTESTCD

Fréquence cardiaque VSTEST VSORRES / minute VSORRESU VSTESTCD

Taille VSTEST VSORRES cm VSORRESU VSTESTCD

Examen clinique VSTEST VSORRES VSTESTCD

Examen neurologique VSTEST VSORRES VSTESTCD

Si anormal, préciser VSTEXT

(VSTEST femmes) Le test urinaire de grossesse est-il négatif VSORRES VSTESTCD

Figure 8 :Aperçu du CO annoté fourni par Ennov Clinical avec un format CDISC

5.1.2.2. Repenser la structure en amont

Ce travail a également mis en lumière une question relative à la création du CRF. Etant donné que seul le standard SDTM a été mis en place dans le cadre de cette réalisation et non pas ceux qui sont en amont pour la définition du CRF (CDASH principalement), il est nécessaire de réfléchir aux tables SDTM dès sa création. Ainsi la question qui se pose est : le CRF doit-il s'adapter au SDTM ou bien faut-il adapter le SDTM au CRF ?

Par exemple la fiche-type d'événements indésirables à l'USMR contient une variable « Grade maximal atteint » qui n'existe pas dans le domaine « *Adverse Events* ». Ce domaine est l'un des domaines qui conservent une structure horizontale, il n'est donc pas possible de rajouter cette variable si l'on souhaite faire du pur CDISC. La solution dans ce cas est de définir un nouvel événement indésirable à chaque fois que le grade change, et donc de pouvoir retrouver le grade maximal atteint.

Il est donc nécessaire de repenser le CRF pour pouvoir appliquer un strict SDTM. En revanche, si l'on souhaite uniquement standardiser les pratiques au sein de l'unité, il est possible de mettre en place un SDTM-like et donc d'ajouter la variable en question dans la table « *Adverse Events* ».

5.1.3. Refonte des circuits de données

L'USMR est aujourd'hui dans une phase de complète refonte des circuits de données. En effet, le service de dactylo-codage de l'université, utilisé pour la saisie des cahiers papiers, est amené à fermer. Pour pallier à cela, le circuit papier a été remplacé par un nouveau circuit électronique avec le recours à un logiciel moins lourd qu'Ennov Clinical : REDCap. Cette migration implique de nombreux changements et nécessite de définir toutes les procédures nécessaires à la mise en place d'une étude sous ce nouvel outil.

Il s'agit là d'un travail conséquent, le choix a donc été fait d'homogénéiser autant que faire se peut les pratiques entre les deux circuits électroniques, pour limiter le nombre de documents qualité à créer. Or REDCap n'est pas nativement compatible avec CDISC, contrairement à Ennov Clinical. En effet, dans REDCap un champ n'est assigné qu'à une seule variable et il n'est donc pas possible de regrouper plusieurs observations sous une même variable comme dans le SDTM.

Pour cette raison, il semble peu probable que le CDISC soit mis en place dans un futur proche pour la gestion des données à l'USMR. Il en va de même pour un standard s'inspirant du CDISC, puisque le principe de verticalité ne pourrait être mis en place sous REDCap pour le moment. Mais les mises à jour du logiciel étant très fréquentes, il est possible que ces possibilités soient développées à l'avenir et que la question de la mise en place du CDISC ou d'un CDISC-like se pose à nouveau à l'USMR. Cependant, d'autres points sont également à prendre en compte. Ils sont développés dans la partie suivante.

5.1.4. Intérêt du CDISC pour l'USMR

La mise en place d'un standard CDISC tel que le SDTM pourrait permettre de standardiser les pratiques pour la gestion des données à l'USMR. En revanche, utilisé seul, il ne permet pas d'exploiter tout l'intérêt des standards. En effet, pour déployer pleinement son potentiel, il faudrait également mettre en place d'autres standards CDISC comme le CDASH dès le développement du cahier d'observation mais surtout ADaM pour la réalisation et la présentation des analyses statistiques. Cela permettrait de ne pas avoir à transformer les tables avant l'analyse et donc d'être plus cohérent dans notre démarche de standardisation.

Cependant, la mise en place d'une standardisation, que ce soit en pur CDISC ou en CDISC-like, nécessite une grande implication du personnel afin de parvenir à des accords sur les définitions des procédures à mettre en œuvre pour la gestion des données et les analyses. D'autant plus que l'unité accompagne aujourd'hui des projets sur des thématiques très variées et qu'il n'est donc pas possible d'utiliser les guides d'implémentation publiés par le CDISC qui ne traite que d'une aire thérapeutique en particulier. Or l'USMR fonctionne actuellement en sous-effectif et n'a pas les moyens nécessaires à cette mise en place.

Concernant la simplification des méta-analyses grâce au recours au CDISC, l'unité est peu sollicitée pour ce type de problématique. Il arrive en revanche que les données d'une étude fassent l'objet d'une extraction *a posteriori* afin que d'autres équipes de recherche en effectuent. La question se pose alors du bénéfice pour l'USMR de déployer tous les efforts nécessaires à la mise en place du CDISC pour des études qui seraient identifiées comme potentiellement pertinentes pour de futures méta-analyses. Cela permettrait aux équipes effectuant ces analyses de ne pas avoir à effectuer un important travail de *data management* avant l'exploitation des données et donc d'arriver plus rapidement à des résultats. Si les moyens étaient donnés à l'unité de le faire, cette question serait alors étudiée.

Au terme de ce travail, la mise en place des standards du CDISC n'a donc pas été identifiée comme nécessaire au vue des pratiques actuelles et des contraintes de l'USMR. En revanche, la standardisation des pratiques au sein de l'équipe l'est, et tout l'intérêt du SMQ est de parvenir à des accords pour définir les procédures à mettre en œuvre. Ces procédures évoluent progressivement et intègrent parfois des principes repris du CDISC, comme par exemple les noms des tables sur deux caractères, ou les noms de variables n'en dépassant pas huit. Il serait peut-être alors intéressant d'intégrer dans cette réflexion la possibilité d'un futur *mapping* CDISC afin de faciliter le travail de personnes qui pourraient souhaiter passer d'un format horizontal à un format CDISC (inversement au programme qui a été développé).

Il faudra également rester informé sur les évolutions des différents standards, et notamment être attentif à la convergence probable de HL7 FHIR et CDISC. En effet, il est possible que dans le futur l'USMR soit amenée à travailler notamment sur des données de l'entrepôt de données du CHU de Bordeaux et il pourrait être intéressant de prendre en compte la possibilité du recours à un standard unique afin de faciliter les échanges de données.

5.2. Performances du programme et limites

Le programme développé permet de transposer les tables d'un format CDISC vers un format horizontal. Cependant, il possède plusieurs limites explicitées ci-après. Il reste donc perfectible

5.2.1. Performances

Afin de juger de l'efficacité de la phase d'optimisation, le temps d'exécution a été calculé par pour le programme avant et après optimisation comme expliqué dans la méthode. Le Tableau 7 présenté dans les résultats page 30 résume les temps d'exécution pour les dérivées 2, 3 et 4. Il s'avère qu'après optimisation du programme, c'est-à-dire après réduction du nombre de boucles et de macros-variables définies par le programme, ce dernier semble être légèrement moins performant qu'avant. Cela pourrait venir du fait que l'opérateur « *like* » est utilisé. A noter que les mesures ont été faites sur les tables contenant les données des patients-tests, c'est-à-dire seulement une trentaine, les temps d'exécution s'allongeraient donc forcément avec des tables plus volumineuses.

Si elle n'a pas permis d'améliorer le critère du temps d'exécution, la phase d'optimisation a toutefois permis d'effectuer des instructions plus cohérentes, qui sont davantage en adéquation avec les données réelles des tables originales. De plus, le programme est amené à n'être utilisé qu'occasionnellement, ce qui ne nécessite pas une importante optimisation du temps d'exécution. C'est pourquoi la deuxième version du programme a tout de même été gardée.

5.2.2. Limites liées au CDISC-*like*

Le fait de ne pas recourir au pur CDISC implique plusieurs adaptations nécessaires à la programmation, en particulier pour l'application des formats. Pour pouvoir réappliquer les formats des variables codées après transposition, deux conditions sont nécessaires. D'une part il faut que les tables contiennent une variable supplémentaire permettant de renseigner le format de l'observation le cas échéant, ce qui a été fait. D'autre part, il faut disposer d'un fichier listant l'ensemble des formats utilisés dans l'eCRF avec toutes les modalités possibles. Cette deuxième contrainte se résout grâce à l'outil de requête proposé par CSExport comme expliqué dans la méthode, mais une manipulation du fichier obtenu est nécessaire. En effet, sous SAS les formats peuvent être appliqués soit à une variable numérique soit à une variable caractère, et dans ce deuxième cas le nom du format doit être précédé d'un « \$ ». Dans le programme développé, les formats s'appliquent aux observations de la variable « --ORRES », qui doivent normalement être au format numérique, mais pour pallier au cas où elles ne le seraient pas (par exemple si une observation est une date), tous les formats du fichier ont été dupliqués pour couvrir à la fois les variables numériques et les variables caractères. Cela permet de pouvoir appliquer les formats quel que soit le format initial de la variable « --ORRES », avec comme limite le respect du cahier des charges qui définit comme condition d'avoir des variables numériques. Cette condition ne serait pas respectée avec le programme actuel si la variable « --ORRES » est initialement au format caractère.

La seconde limite liée au CDISC-*like* relève du paramétrage. En effet, le choix a été fait de ne pas garder uniquement « --ORRES » et « --TEST » comme variable dans le recours au CDISC-*like*. Ainsi, les tables contiennent parfois d'autres variables permettant le renseignement de l'unité, une précision ou un champ calculé par exemple. De ce fait, il n'est pas possible d'effectuer un paramétrage automatique permettant de faire tourner le programme sur les mêmes variables pour toutes les tables. Il est donc nécessaire de paramétrer le programme manuellement en renseignant, pour chaque table, les variables CDISC que l'on souhaite transposer. Ce paramétrage devient donc d'autant plus fastidieux que le nombre de tables à traiter est important. Cela peut par nature être source d'erreur et donc engendrer le mauvais déroulement du programme. Pour résoudre en partie cette difficulté, le rapport PDF édité en fin de programme notifie toutes les erreurs remontées dans la *log*. Mais il est possible qu'une erreur n'empêche pas le programme de tourner et dans ce cas le résultat obtenu ne sera pas celui escompté. Il est donc important de bien se référer à la notice d'utilisation du programme pour remplir correctement les paramètres.

5.2.3. Autres limites identifiées

Le programme développé contient une autre limite concernant la suppression de certaines colonnes. En effet, le programme permet de supprimer les colonnes ne contenant aucune donnée dans les tables transposées horizontales ou semi-horizontales (contenant les visites les unes au-dessous des autres). Cela permet de supprimer les colonnes créées par le programme alors qu'elles n'existent pas dans les données originales, comme par exemple la variable concernant l'unité du résultat de l'examen clinique (il s'agit d'une variable codée « normal » ou « anormal », elle n'a donc pas d'unité). La création de ces variables s'explique par le fait que, pour chaque variable CDISC que l'on souhaite transposer, une colonne est créée pour chaque observation dans la table finale. Ainsi, si la table originale contient trois variables CDISC que l'on souhaite transposer et dix observations, la table transposée contiendra trente colonnes (en plus des variables identifiantes comme « USUBJID » ou « DOMAIN »), même si les variables CDISC ne sont pas pertinentes pour l'ensemble des observations. Il est donc nécessaire de supprimer par la suite ces colonnes afin de ne garder que les variables ayant du sens. Or, il est possible qu'une variable ait du sens mais qu'elle ne soit renseignée pour aucun participant. Cette variable devrait normalement être dans la table finale. Mais le programme tel qu'il est construit ne fait pas la différence entre ces deux cas et supprime quoi qu'il arrive les colonnes vides. Un message d'avertissement a donc été ajouté dans la notice du programme pour informer l'utilisateur.

5.2.4. Pistes d'amélioration

Dans un souci d'ergonomie, le programme développé pourrait être amélioré par l'ajout d'une interface. Cela a été envisagé à l'aide d'un module appelé SAS/AF mais sa non disponibilité au sein de l'USMR n'a pas permis son application. Il aurait également été possible d'ouvrir une fenêtre Windows sous SAS à l'aide de la commande %window comme sur l'exemple de la Figure 9. Cela permet à l'utilisateur de remplir les paramètres nécessaires directement dans la fenêtre et de les réinjecter dans le programme ensuite. Cependant le nombre de tables à

transposer et le nombre de variables pour chaque table n'étant pas fixes, son implémentation n'a pas pu être mise en place.

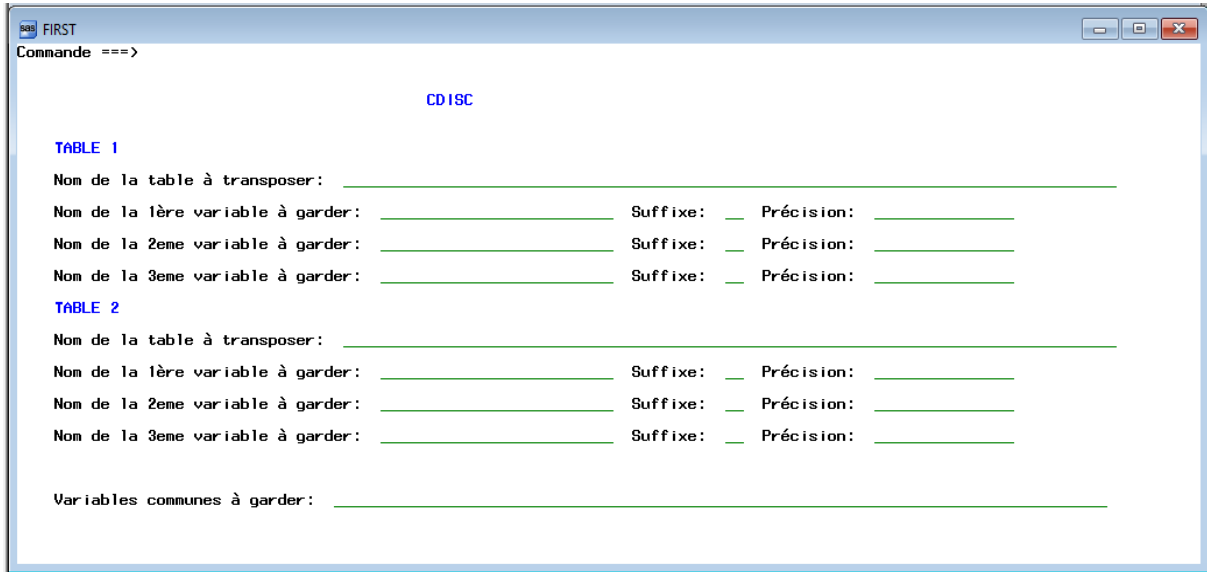


Figure 9 : Exemple de fenêtre Windows affichée par SAS

Le recours à de la programmation en VBA (*Visual Basic for Applications*) sous Excel pourrait permettre, à partir du remplissage d'un tableur, de créer une sorte d'interface qui permettrait à l'utilisateur de choisir quelle dérivée et quel(s) module(s) il souhaite lancer. Si le programme Croche-patte est amené à être utilisé régulièrement, cette solution pourrait être envisagée par la suite.

6. CONCLUSION

La gestion des données nécessite une standardisation des pratiques afin de pouvoir assurer la fiabilité et la qualité des données tout en réduisant les temps de traitement et donc les coûts. Cette nécessité est d'autant plus grande qu'un courant se renforce afin de rendre les données plus accessibles et plus facilement réutilisables. Cela amène à l'application de standards qui permettent l'interopérabilité entre les systèmes et donc facilitent les échanges de données.

Ce travail a permis de mettre en lumière l'intérêt que peuvent avoir les standards pour la recherche. Cependant, différentes difficultés ont été rencontrées : la complexité de compréhension et de mise en place d'un standard strict, le temps nécessaire à cette mise en place et le manque de compatibilité de certains logiciels de gestion de données (REDCap par exemple). De plus, au regard des activités actuelles de l'USMR, la mise en place d'un standard tel que le CDISC n'a pas été jugée nécessaire. La conclusion de ce travail est donc que, bien qu'il soit faisable d'appliquer le CDISC à l'USMR si les moyens nécessaires étaient donnés à l'équipe, il n'a pas été décidé de le mettre en place pour le moment.

En revanche, le programme développé dans le cadre de ce projet et en cours de validation conformément au SMQ pourra être utilisé soit par les gestionnaires de données, soit par les statisticiens dans le cas où des données externes seraient récoltées pour une étude et envoyées dans un format CDISC. Si son utilisation est amenée à devenir fréquente, des améliorations telles que l'ajout d'une interface pourraient être envisagées.

Travailler au sein de l'USMR a permis de mettre en œuvre différentes connaissances et compétences acquises lors du master : la culture générale relative à la recherche clinique assimilée lors des deux années de formation, l'initiation au logiciel SAS effectuée lors de la première année et la sensibilisation à la thématique de l'interopérabilité en santé et des standards lors de la deuxième année. Les connaissances acquises en matière de base de données ont également été utiles, en particulier le langage SQL.

7. REFERENCES

1. Unité de Soutien Méthodologique à la Recherche clinique et épidémiologique (USMR). USMR [Internet]. 2017 [cited 2019 Feb 25]. Available from: <http://usmr.isped.u-bordeaux2.fr/index.asp?act=1>
2. Les Entreprises du Médicament (LEEM). Atelier presse ‘Essais cliniques’ - Les études cliniques en 20 questions [Internet]. 2011 [cited 2019 Apr 4]. Available from: <https://www.leem.org/sites/default/files/EtUDES%20CLINIQUES20questionsmai200%5B1%5D.pdf>
3. Yamamoto K, Ota K, Akiya I, Shintani A. A pragmatic method for transforming clinical research data from the research electronic data capture “REDCap” to Clinical Data Interchange Standards Consortium (CDISC) Study Data Tabulation Model (SDTM): Development and evaluation of REDCap2SDTM. *J Biomed Inform.* 2017;70:65–76.
4. Krishnankutty B, Bellary S, Kumar NBR, Moodahadu LS. Data management in clinical research: An overview. *Indian J Pharmacol.* 2012;44(2):168–72.
5. Organisation de Coopération et de Développement Economiques (OCDE). Data standardization Definition [Internet]. [cited 2019 Mar 25]. Available from: <https://stats.oecd.org/glossary/detail.asp?ID=4501>
6. Hudson LD, Kush RD, Almario EN, Seigneuret N, Jackson T, Jauregui B, et al. Global Standards to Expedite Learning From Medical Research Data. *Clin Transl Sci.* 2018;11(4):342–4.
7. Huser V, Sastry C, Breymaier M, Idriss A, Cimino JJ. Standardizing data exchange for clinical research protocols and case report forms: An assessment of the suitability of the Clinical Data Interchange Standards Consortium (CDISC) Operational Data Model (ODM). *J Biomed Inform.* 2015;57:88–99.
8. TICsanté. L’ASIP Santé publie la version 1.0 du cadre d’interopérabilité des systèmes d’information de santé [Internet]. [cited 2019 Feb 19]. Available from: <https://www.ticsante.com/story/785/l’asip-sante-publie-la-version-10-du-cadre-d’interopabilite-des-systemes-d’information-de-sante.html>
9. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data.* 2016;3:160018.
10. Taichman DB, Sahni P, Pinborg A, Peiperl L, Laine C, James A, et al. Data Sharing Statements for Clinical Trials: A Requirement of the International Committee of Medical Journal Editors. *Ethiop J Health Sci.* 2017;27(4):315–8.
11. Clinical Data Interchange Standards Consortium (CDISC) Portal. CDISC : introduction générale [Internet]. [cited 2018 Oct 30]. Available from: <http://portal.cdisc.org/CDISC%20User%20Networks/Europe/French%20Language/Prsentation%20des%20Standards/Home.aspx>
12. Hume S, Aerts J, Sarnikar S, Huser V. Current applications and future directions for the CDISC Operational Data Model standard: A methodological review. *J Biomed Inform.* 2016 Apr;60:352–62.

13. CDISC Portal. Clinical Data Acquisition Standards Harmonization (CDASH) [Internet]. [cited 2018 Oct 30]. Available from: [http://portal.cdisc.org/CDISC%20User%20Networks/Europe/French%20Language/CDA SH1/Home.aspx](http://portal.cdisc.org/CDISC%20User%20Networks/Europe/French%20Language/CDA%20SH1/Home.aspx)
14. CDISC Portal. Study Data Tabulation Model (SDTM) [Internet]. [cited 2018 Oct 30]. Available from: <http://portal.cdisc.org/CDISC%20User%20Networks/Europe/French%20Language/SDTM/Home.aspx>
15. CDISC. Therapeutic Area - Published User Guides [Internet]. CDISC. [cited 2019 Feb 20]. Available from: <https://www.cdisc.org/standards/therapeutic-areas>
16. WorldWide Antimalarial Resistance Network. La puissance de la normalisation: élaboration de normes CDISC pour le paludisme [Internet]. 2015 [cited 2019 Feb 19]. Available from: <https://www.wwarn.org/fr/actualite/la-puissance-de-la-normalisation-elaboration-de-normes-cdisc-pour-le-paludisme>
17. CDISC. Use Cases for Clear Data [Internet]. CDISC. [cited 2018 Oct 30]. Available from: <https://www.cdisc.org/use-cases-for-clear-data>
18. Matsumura Y, Hattori A, Manabe S, Takeda T, Takahashi D, Yamamoto Y, et al. Interconnection of electronic medical record with clinical data management system by CDISC ODM. *Stud Health Technol Inform.* 2014;205:868–72.
19. Chateau Y, Schiappa R, Gal J, Viotti J, Bozec A, Etienne-Grimaldi M-C, et al. Entrepôt de données des essais thérapeutiques, valoriser quinze ans de données de recherche clinique. *Rev Epidemiol Santé Publique.* 2018;66:S133–4.
20. Roca L, Boudon S, Juzyn B, Gourgou S, Martin A-L. Data Center UNICANCER R&D. *Rev Epidemiol Santé Publique.* 2014;62:S137.
21. Sato I, Kawasaki Y, Ide K, Sakakibara I, Konomura K, Yamada H, et al. Clinical Data Interchange Standards Consortium Standardization of Biobank Data: A Feasibility Study. *Biopreserv Biobank.* 2016;14(1):45–50.
22. Health Level 7 (HL7). Introducing HL7 Fast Healthcare Interoperability Resources (FHIR) [Internet]. [cited 2019 Feb 19]. Available from: <http://hl7.org/fhir/summary.html>
23. Leroux H, Metke-Jimenez A, Lawley MJ. Towards achieving semantic interoperability of clinical study data with FHIR. *J Biomed Semant.* 2017;8(1).
24. Aerts J. Towards a Single Data Exchange Standard for Use in Healthcare and in Clinical Research. *Stud Health Technol Inform.* 2018;248:55–63.
25. Zopf R, Abolafia J, Reddy B. Use of Fast Healthcare Interoperability Resources (FHIR) in the Generation of Real World Evidence (RWE) [Internet]. PhUSE Annual Conference; 2017 Oct 11 [cited 2019 Feb 19]; Edinburgh. Available from: https://www.researchgate.net/publication/322675315_Use_of_Fast_Healthcare_Interoperability_Resources_FHIR_in_the_Generation_of_Real_World_Evidence_RWE
26. Groupe des Data Managers Académiques (acaDM). Traduction CDASH et SDTM [Internet]. [cited 2018 Nov 28]. Available from: <http://www.acadm.fr/pages/yacacdisc/traduction-des-variables-cdash-et-sdtm.html>

8. ANNEXES

Table des annexes :

ANNEXE 1: Domaines du SDTM	48
ANNEXE 2: SDTM créé pour l'étude SOSrial	49
ANNEXE 3 : Variables par domaine du SDTM créé pour SOSrial	53
ANNEXE 4 : Paramétrage des tables sous CSExport.....	56
ANNEXE 5 : Cahier des charges pour des tables fournies pour l'analyse	59
ANNEXE 6 : Code du programme Croche-patte.....	60
ANNEXE 7 : Notice d'utilisation du programme Croche-patte	71
ANNEXE 8 : Protocole de qualification du programme Croche-patte	72
ANNEXE 9 : Rapport de qualification du programme Croche-patte	73
ANNEXE 10 : Script de l'intervention en forum méthodologique du 12/06/2019	75

ANNEXE 1: Domaines du SDTM

CLASS	DOMAIN NAME	DOMAIN DESCRIPTION
Special purpose	CO	Comments
	DM	Demographics
	SE	Subject Elements
	SV	Subject Visits
Interventions	CM	Concomitant Medications
	EC	Exposure as Collected
	EX	Exposure
	PR	Procedures
	SU	Substance Use
Events	AE	Adverse Events
	CE	Clinical Events
	DS	Disposition
	DV	Protocol Deviation
	HO	Health Encounters
	MH	Medical History
Findings	DA	Drug Accountability
	DD	Death Details
	EG	ECG Test Results
	IE	Inclusion/Exclusion criterion not Met
	IS	Immunogenicity Specimen Assessments
	LB	Laboratory Test Results
	MB	Microbiology Specimen
	MI	Microscopic Findings
	MO	Morphology
	MS	Microbiology Susceptibility Test
	PC	PK Concentrations
	PE	Physical Examination
	PP	PK Parameters
	QS	Questionnaires
	RP	Reproductive System Findings
	RS	Disease Response
	SC	Subject Characteristics
	SS	Subject Status
	TR	Tumor Results
	TU	Tumor Identification
VS	Vital Signs	
Findings about	FA	Findings About
	SR	Skin Response
Trial Design Domains	TA	Trial Arms
	TD	Trial Disease Assessment
	TE	Trial Elements
	TI	Trial Inclusion/Exclusion Criteria
	TS	Trial Summary
	TV	Trial Visits
Relationship Datasets	SUPP--	Supplemental Qualifiers
	RELREC	Related Records

ANNEXE 2: SDTM créé pour l'étude SOStrial

Domaine SV (« Subject Visit »)

STUDYID	DOMAIN	USUBJID	VISITNUM	VISIT	SVSTAT	SVSDTC	SVREATEL	SVREASND	SVACCOMP
SOSTRIAL	SV		1	1 PRE-INCLUSION	fait sur site	13/07/2017			
SOSTRIAL	SV		1	2 INCLUSION	fait sur site	20/07/2017			participant
SOSTRIAL	SV		1	3 T3M	fait par téléphone	21/10/2017	Raison médicale		aidant
SOSTRIAL	SV		1	4 T6M/RANDO					
SOSTRIAL	SV		1	5 T12M					
SOSTRIAL	SV		1	6 T18M					
SOSTRIAL	SV		1	7 T24M					
SOSTRIAL	SV		1	8 T30M					

Domaine DM (« Demographics »)

STUDYID	DOMAIN	USUBJID	BRTHDTC	SEX	ARM	ARMCD	RANDONAM	RANDODTC	RFSTDTCT	RFENDTCT	RFXSTDTCT	RFXENDTCT	RFICDTC	DTHFL	DTHDTC
SOSTRIAL	DM		1	16/06/1948	F				début suivi	fin suivi	premier trait	dernier trait	13/07/2017		0

Domaine SC (« Subject Characteristics »)

STUDYID	DOMAIN	USUBJID	VISITNUM	VISIT	SCSEQ	SCTESTCD	SCTEST	SCORRES	SCSTRESC
SOSTRIAL	SC		1	1 PRE-INCLUSION	1	ALD	Le patient bénéficie-t-il de l'ALD ?		1 Oui
SOSTRIAL	SC		1	1 PRE-INCLUSION	2	SOCIOEDU	Niveau socio-éducatif du participant		
SOSTRIAL	SC		1	1 PRE-INCLUSION	3	SITPRO	Situation professionnelle actuelle		Retraité
SOSTRIAL	SC		1	1 PRE-INCLUSION	4	DERNPRO	Si SITPRO = chômeur, retraité, invalide ou congé longue maladie, dernière profession exercée		Cadre
SOSTRIAL	SC		1	1 PRE-INCLUSION	5	ACTPROYN	Si SITPRO = h/f au foyer ou autre (rentier, bénévole), le participant a-t-il déjà exercé une activité pro ?		
SOSTRIAL	SC		1	1 PRE-INCLUSION	6	PROYPRE	Si ACTPROYN = oui, dernière profession exercée		
SOSTRIAL	SC		1	1 PRE-INCLUSION	7	AGESTOP	Si ACTPROYN = oui, age lors de la cessation d'activité		
SOSTRIAL	SC		1	1 PRE-INCLUSION	8	PROLONG	Si différent de SITPRO, profession exercée le plus longtemps		
SOSTRIAL	SC		1	1 PRE-INCLUSION	9	PROLONGD	Durée de l'activité exercée le plus longtemps	25	25
SOSTRIAL	SC		1	1 PRE-INCLUSION	10	MARISTAT	Etat civil actuel du participant		1 Marié
SOSTRIAL	SC		1	1 PRE-INCLUSION	11	DIVSEPY	Si MARISTAT = divorcé ou séparé, année de divorce ou de séparation		
SOSTRIAL	SC		1	1 PRE-INCLUSION	12	DCCONJY	Si MARISTAT = veuf ou décès du conjoint, année de décès du conjoint		
SOSTRIAL	SC		1	1 PRE-INCLUSION	13	MARCONCY	Si MARISTAT = marié ou concubin, année de mariage ou concubinage	1969	1969
SOSTRIAL	SC		1	1 PRE-INCLUSION	14	ACCOMYN	Le participant est-il venu accompagné à la visite de pré-inclusion ?		1 Oui
SOSTRIAL	SC		1	1 PRE-INCLUSION	15	ACCOMPR	Si ACCOMYN = Oui, avec qui le participant est-il venu ?		Fille/fils
SOSTRIAL	SC		1	1 PRE-INCLUSION	16	ACCOMPR	Si ACCOMPR = autre, précision		
SOSTRIAL	SC		1	1 PRE-INCLUSION	17	VITSEUL	Le participant vit-il seul ?		0 Non
SOSTRIAL	SC		1	1 PRE-INCLUSION	18	VITCONJ	Si VITSEUL = non, vit-il avec son conjoint ou concubin ?		1 Oui
SOSTRIAL	SC		1	1 PRE-INCLUSION	19	VITENF	Si VITSEUL = non, vit-il avec son/ses enfant(s) ?		0 Non
SOSTRIAL	SC		1	1 PRE-INCLUSION	20	VITFRERE	Si VITSEUL = non, vit-il avec son/ses frere(s) ou sœur(s) ?		0 Non
SOSTRIAL	SC		1	1 PRE-INCLUSION	21	VITPAR	Si VITSEUL = non, vit-il avec son père ou sa mère ?		0 Non
SOSTRIAL	SC		1	1 PRE-INCLUSION	22	VITCOHAB	Si VITSEUL = non, vit-il en cohabitation non familiale (ami, autre) ?		0 Non
SOSTRIAL	SC		1	1 PRE-INCLUSION	23	TUTEL	Le participant est-il sous tutelle ?		0 Non
SOSTRIAL	SC		1	1 PRE-INCLUSION	24	CURATEL	Le participant est-il sous curatelle ?		0 Non
SOSTRIAL	SC		1	1 PRE-INCLUSION	25	DEMENCE	Le participant présente-t-il une démence ?		1 Liée à la maladie d'Alzheimer
SOSTRIAL	SC		1	1 PRE-INCLUSION	26	DATESYMP	Date des premiers symptômes	12/06/2016	12/06/2016
SOSTRIAL	SC		1	1 PRE-INCLUSION	27	DATEDIAG	Date du diagnostic de la maladie d'Alzheimer selon les critères NINCDS-ARDRA	23/05/2017	26/05/2017
SOSTRIAL	SC		1	1 PRE-INCLUSION	28	ATCDPAT	Le participant a-t-il des antécédents ou pathologies ?		1 Oui
SOSTRIAL	SC		1	1 PRE-INCLUSION	29	TTYN	Le participant prend-il des médicaments ?		0 Non

Domaine EX (« Exposure »)

STUDYID	DOMAIN	USUBJID	VISITNUM	VISIT	EXSEQ	EXCAT	EXTRT	EXDOSE	EXDOSU	EXDOSFRM	EXDOSTOT	EXADJ	EXSTDTCT	EXENDTCT
SOSTRIAL	EX		1	1 PRE-INCLUSION		1	IAC	abc	1 mg	comprimé		4 raison ajustement		

Domaine TI (« Trial Inclusion/Exclusion criteria »)

STUDYID	DOMAIN	USUBJID	VISITNUM	VISIT	TISEQ	TITESTCD	TITEST	TIORRES	TISTRESC	TIDTC
SOSTRIAL	TI		1	1 PRE-INCLUSION	1	INCL01	Nouveau cas de maladie d'Alzheimer consultant dans un CMRR ou une CM		1 Oui	
SOSTRIAL	TI		1	1 PRE-INCLUSION	2	INCL02	Diagnostic probable ou possible MA selon NINCDS-ARDRA		1 Oui	
SOSTRIAL	TI		1	1 PRE-INCLUSION	3	INCL03	Stade léger à modéré (MMSE>15 à la pré-inclusion)		1 Oui	
SOSTRIAL	TI		1	1 PRE-INCLUSION	4	INCL04	Patient avec une indication pour un traitement par IAC		1 Oui	
SOSTRIAL	TI		1	1 PRE-INCLUSION	5	INCL05	Patient âgé de 50 ans ou plus		1 Oui	
SOSTRIAL	TI		1	1 PRE-INCLUSION	6	INCL06	Si SEX=F, patiente ménopausée ou ayant une contraception efficace		1 Oui	
SOSTRIAL	TI		1	1 PRE-INCLUSION	7	INCL07	Personne affiliée ou bénéficiaire d'un régime de SS		1 Oui	
SOSTRIAL	TI		1	1 PRE-INCLUSION	8	INCL08	Patient bénéficiant de l'ALD liée à la MA		1 Oui	
SOSTRIAL	TI		1	1 PRE-INCLUSION	9	INCL09	Signature du consentement éclairé par le participant et son aidant		1 Oui	
SOSTRIAL	TI		1	1 PRE-INCLUSION	10	EXCL01	Démence d'une autre étiologie que la MA		0 Non	
SOSTRIAL	TI		1	1 PRE-INCLUSION	11	EXCL02	MA avec un MMSE <=15 à la pré-inclusion		0 Non	
SOSTRIAL	TI		1	1 PRE-INCLUSION	12	EXCL03	Patient présentant une contre-indication aux IAC		0 Non	
SOSTRIAL	TI		1	1 PRE-INCLUSION	13	EXCL04	Patient résidant en institution au moment de l'inclusion		0 Non	
SOSTRIAL	TI		1	1 PRE-INCLUSION	14	EXCL05	Patient présentant une dépendance complète pour la toilette ou l'habillement à l'inclusion		0 Non	
SOSTRIAL	TI		1	1 PRE-INCLUSION	15	EXCL06	Patient sous tutelle ou curatelle, patient incapable d'exprimer son consentement		0 Non	
SOSTRIAL	TI		1	1 PRE-INCLUSION	16	EXCL07	Patient avec une comorbidité sévère ou instable compromettant son suivi dans l'étude		0 Non	
SOSTRIAL	TI		1	1 PRE-INCLUSION	17	EXCL08	Patient sans aidant identifiable		0 Non	
SOSTRIAL	TI		1	1 PRE-INCLUSION	18	EXCL09	Patient déjà inclus dans un autre protocole d'essai thérapeutique		0 Non	
SOSTRIAL	TI		1	1 PRE-INCLUSION	19	EXCL10	Si SEX=F, patiente enceinte ou allaitante		0 Non	
SOSTRIAL	TI		1	4 T6M/RANDO	20	EXCL01	Patient répondeur au traitement par IAC avec MMSE stable ou augmentant après 6 mois de traitement			
SOSTRIAL	TI		1	4 T6M/RANDO	21	EXCL02	Patient avec dépendance totale à la toilette ou à l'habillement au moment de la randomisation			
SOSTRIAL	TI		1	4 T6M/RANDO	22	EXCL03	Patient résidant en institution au moment de la randomisation			

Domaine QS (« Questionnaires »). Dans un souci de lisibilité il a été tronquée (seule la visite 1 apparaît ici alors que les mêmes lignes sont répétées pour les visites 4, 5, 6, 7 et 8)

STUDYID	DOMAIN	USUBJID	VISITNUM	VISIT	OSSEQ	QSDTC	QSCAT	QSTESTCD	QSTESTCD2	QSTEST	QSORRES	QSTRESC	QSTAT	QREASND
SOSTRIAL	QS		1	1 PRE-INCLUSION	1	13/07/2017	ADL-KATZ	DEPTOIL	ADJ01	Dépendance de la toilette (échelle ADL de KATZ)		0	besoin d'aucune aide	
SOSTRIAL	QS		1	1 PRE-INCLUSION	2	13/07/2017	ADL-KATZ	DEPHAB	ADJ02	Dépendance de l'habillement (échelle ADL de KATZ)		0	besoin d'aucune aide	
SOSTRIAL	QS		1	1 PRE-INCLUSION	3	13/07/2017	ADL-KATZ	DEPWC	ADJ03	Dépendance pour les WC (échelle ADL de KATZ)		1	besoin d'une aide	
SOSTRIAL	QS		1	1 PRE-INCLUSION	4	13/07/2017	ADL-KATZ	DEPLOCOM	ADJ04	Dépendance de la locomotion (échelle ADL de KATZ)		1	besoin d'une aide	
SOSTRIAL	QS		1	1 PRE-INCLUSION	5	13/07/2017	ADL-KATZ	DEPALIM	ADJ05	Dépendance de l'alimentation (échelle ADL de KATZ)		1	besoin d'aide pour couper la viande ou beurrer le pain	
SOSTRIAL	QS		1	1 PRE-INCLUSION	6	13/07/2017	ADL-KATZ	ADLSOCOR	ADLSCOT	Score total de l'échelle ADL de Katz		3		3
SOSTRIAL	QS		1	1 PRE-INCLUSION	7	13/07/2017	MMSE	OTANNEE	MMSE01	En quelle année sommes-nous?		1		1
SOSTRIAL	QS		1	1 PRE-INCLUSION	8	13/07/2017	MMSE	OTANNEE	MMSE02	En quelle saison sommes-nous?		0		0
SOSTRIAL	QS		1	1 PRE-INCLUSION	9	13/07/2017	MMSE	OTMOIS	MMSE03	En quel mois sommes-nous?		1		1
SOSTRIAL	QS		1	1 PRE-INCLUSION	10	13/07/2017	MMSE	OTJOURM	MMSE04	Quel jour du mois sommes-nous?		0		0
SOSTRIAL	QS		1	1 PRE-INCLUSION	11	13/07/2017	MMSE	OTJOURS	MMSE05	Quel jour de la semaine sommes-nous?		1		1
SOSTRIAL	QS		1	1 PRE-INCLUSION	12	13/07/2017	MMSE	OEHOP	MMSE06	Quel est le nom de l'hôpital?		0		0
SOSTRIAL	QS		1	1 PRE-INCLUSION	13	13/07/2017	MMSE	OEVILLE	MMSE07	Dans quelle ville sommes-nous?		1		1
SOSTRIAL	QS		1	1 PRE-INCLUSION	14	13/07/2017	MMSE	OEEPT	MMSE08	Dans quel département sommes-nous?		1		1
SOSTRIAL	QS		1	1 PRE-INCLUSION	15	13/07/2017	MMSE	OEREG	MMSE09	Dans quelle région sommes-nous?		0		0
SOSTRIAL	QS		1	1 PRE-INCLUSION	16	13/07/2017	MMSE	OETAG	MMSE10	A quel étage sommes-nous?		0		0
SOSTRIAL	QS		1	1 PRE-INCLUSION	17	13/07/2017	MMSE	SCOREOET	MMSESCO1	Score total d'orientation espace-temps		5		5
SOSTRIAL	QS		1	1 PRE-INCLUSION	18	13/07/2017	MMSE	LISTMEMO	LIMMSE	Liste utilisée pour le test de mémorisation		2	Liste 2 (Fleur, clé, tulipe)	
SOSTRIAL	QS		1	1 PRE-INCLUSION	19	13/07/2017	MMSE	SCOREAPP	MMSESCO2	Score apprentissage (nombre de mots retenus)		2		2
SOSTRIAL	QS		1	1 PRE-INCLUSION	20	13/07/2017	MMSE	SOLUST93	MMSE11	Soustraction par 7 première étape		1		1
SOSTRIAL	QS		1	1 PRE-INCLUSION	21	13/07/2017	MMSE	SOLUS86	MMSE12	Soustraction par 7 deuxième étape		1		1
SOSTRIAL	QS		1	1 PRE-INCLUSION	22	13/07/2017	MMSE	SOLUS79	MMSE13	Soustraction par 7 troisième étape		1		1
SOSTRIAL	QS		1	1 PRE-INCLUSION	23	13/07/2017	MMSE	SOLUS72	MMSE14	Soustraction par 7 quatrième étape		0		0
SOSTRIAL	QS		1	1 PRE-INCLUSION	24	13/07/2017	MMSE	SOLUS65	MMSE15	Soustraction par 7 cinquième étape		1		1
SOSTRIAL	QS		1	1 PRE-INCLUSION	25	13/07/2017	MMSE	SCORESOU	MMSESCO3	Score total pour la soustraction par 7		4		4
SOSTRIAL	QS		1	1 PRE-INCLUSION	26	13/07/2017	MMSE	SCORERAP	MMSESCO4	Score rappel (nombre de mots retenus)		1		1
SOSTRIAL	QS		1	1 PRE-INCLUSION	27	13/07/2017	MMSE	LANGCRAY	MMSE16	Désignation d'un crayon		1		1
SOSTRIAL	QS		1	1 PRE-INCLUSION	28	13/07/2017	MMSE	LANGMON	MMSE17	Désignation d'une montre		1		1
SOSTRIAL	QS		1	1 PRE-INCLUSION	29	13/07/2017	MMSE	LANGREP	MMSE18	Répétition d'une phrase		0		0
SOSTRIAL	QS		1	1 PRE-INCLUSION	30	13/07/2017	MMSE	LANGPRE	MMSE19	Préhension d'une feuille de la main droite		1		1
SOSTRIAL	QS		1	1 PRE-INCLUSION	31	13/07/2017	MMSE	LANGPLU	MMSE20	Pliage de la feuille en deux		1		1
SOSTRIAL	QS		1	1 PRE-INCLUSION	32	13/07/2017	MMSE	LANGJET	MMSE21	Jet de la feuille par terre		0		0
SOSTRIAL	QS		1	1 PRE-INCLUSION	33	13/07/2017	MMSE	LANGLIR	MMSE22	Faire ce qu'il est écrit sur la feuille		1		1
SOSTRIAL	QS		1	1 PRE-INCLUSION	34	13/07/2017	MMSE	LANGECR	MMSE23	Ecrire une phrase entière		0		0
SOSTRIAL	QS		1	1 PRE-INCLUSION	35	13/07/2017	MMSE	SCORELAN	MMSESCO5	Score total pour le langage		5		5
SOSTRIAL	QS		1	1 PRE-INCLUSION	36	13/07/2017	MMSE	SCOREDES	MMSESCO6	Score pour le dessin		1		1
SOSTRIAL	QS		1	1 PRE-INCLUSION	37	13/07/2017	MMSE	SCOREMMSE	MMSESCOT	Score total MMSE		18		18

Domaine SS (« Subject Status »)

STUDYID	DOMAIN	USUBJID	VISITNUM	VISIT	SSSEQ	SSTESTCD	SSTEST	SSORRES	SSSTRESC	SSSTAT	SSREASND	SSDTC
SOSTRIAL	SS		1	1 PRE-INCLUSION	1	INSTITUT	Le participant est-il en institution?		0	Non		
SOSTRIAL	SS		1	2 INCLUSION	2	EVESANT	Le participant a-t-il eu un ou plusieurs événement(s) de santé depuis la dernière visite?		0	Non		
SOSTRIAL	SS		1	2 INCLUSION	3	CHGTTT	Le participant a-t-il eu des changements de traitements depuis la dernière visite?		1	Oui		
SOSTRIAL	SS		1	2 INCLUSION	4	HOSPIT	Depuis le dernier suivi, le participant a-t-il eu au moins une hospitalisation?		0	Non		
SOSTRIAL	SS		1	2 INCLUSION	5	DC	Le participant est-il décédé?		0	Non		
SOSTRIAL	SS		1	3 T3M	6	EVESANT	Le participant a-t-il eu un ou plusieurs événement(s) de santé depuis la dernière visite?					
SOSTRIAL	SS		1	3 T3M	7	CHGTTT	Le participant a-t-il eu des changements de traitements depuis la dernière visite?					
SOSTRIAL	SS		1	3 T3M	8	HOSPIT	Depuis le dernier suivi, le participant a-t-il eu au moins une hospitalisation?					
SOSTRIAL	SS		1	3 T3M	9	DC	Le participant est-il décédé?					
SOSTRIAL	SS		1	4 T6M/RANDO	10	CARNPAT	Le carnet patient a-t-il été complété depuis la dernière visite?					
SOSTRIAL	SS		1	4 T6M/RANDO	11	CARNPATN	Si CARNPAT=Non, raison					
SOSTRIAL	SS		1	4 T6M/RANDO	12	EVESANT	Le participant a-t-il eu un ou plusieurs événement(s) de santé depuis la dernière visite?					
SOSTRIAL	SS		1	4 T6M/RANDO	13	CHGTTT	Le participant a-t-il eu des changements de traitements depuis la dernière visite?					
SOSTRIAL	SS		1	4 T6M/RANDO	14	HOSPIT	Depuis le dernier suivi, le participant a-t-il eu au moins une hospitalisation?					
SOSTRIAL	SS		1	4 T6M/RANDO	15	DC	Le participant est-il décédé?					
SOSTRIAL	SS		1	4 T6M/RANDO	16	INSTITUT	Le participant est-il en institution?					
SOSTRIAL	SS		1	5 T12M	17	CARNPAT	Le carnet patient a-t-il été complété depuis la dernière visite?					
SOSTRIAL	SS		1	5 T12M	18	CARNPATN	Si CARNPAT=Non, raison					
SOSTRIAL	SS		1	5 T12M	19	EVESANT	Le participant a-t-il eu un ou plusieurs événement(s) de santé depuis la dernière visite?					
SOSTRIAL	SS		1	5 T12M	20	CHGTTT	Le participant a-t-il eu des changements de traitements depuis la dernière visite?					
SOSTRIAL	SS		1	5 T12M	21	HOSPIT	Depuis le dernier suivi, le participant a-t-il eu au moins une hospitalisation?					
SOSTRIAL	SS		1	5 T12M	22	DC	Le participant est-il décédé?					
SOSTRIAL	SS		1	5 T12M	23	INSTITUT	Le participant est-il en institution?					
SOSTRIAL	SS		1	6 T18M	24	CARNPAT	Le carnet patient a-t-il été complété depuis la dernière visite?					
SOSTRIAL	SS		1	6 T18M	25	CARNPATN	Si CARNPAT=Non, raison					
SOSTRIAL	SS		1	6 T18M	26	EVESANT	Le participant a-t-il eu un ou plusieurs événement(s) de santé depuis la dernière visite?					
SOSTRIAL	SS		1	6 T18M	27	CHGTTT	Le participant a-t-il eu des changements de traitements depuis la dernière visite?					
SOSTRIAL	SS		1	6 T18M	28	HOSPIT	Depuis le dernier suivi, le participant a-t-il eu au moins une hospitalisation?					
SOSTRIAL	SS		1	6 T18M	29	DC	Le participant est-il décédé?					
SOSTRIAL	SS		1	6 T18M	30	INSTITUT	Le participant est-il en institution?					
SOSTRIAL	SS		1	7 T24M	31	CARNPAT	Le carnet patient a-t-il été complété depuis la dernière visite?					
SOSTRIAL	SS		1	7 T24M	32	CARNPATN	Si CARNPAT=Non, raison					
SOSTRIAL	SS		1	7 T24M	33	EVESANT	Le participant a-t-il eu un ou plusieurs événement(s) de santé depuis la dernière visite?					
SOSTRIAL	SS		1	7 T24M	34	CHGTTT	Le participant a-t-il eu des changements de traitements depuis la dernière visite?					
SOSTRIAL	SS		1	7 T24M	35	HOSPIT	Depuis le dernier suivi, le participant a-t-il eu au moins une hospitalisation?					
SOSTRIAL	SS		1	7 T24M	36	DC	Le participant est-il décédé?					
SOSTRIAL	SS		1	7 T24M	37	INSTITUT	Le participant est-il en institution?					
SOSTRIAL	SS		1	8 T30M	38	CARNPAT	Le carnet patient a-t-il été complété depuis la dernière visite?					
SOSTRIAL	SS		1	8 T30M	39	CARNPATN	Si CARNPAT=Non, raison					
SOSTRIAL	SS		1	8 T30M	40	EVESANT	Le participant a-t-il eu un ou plusieurs événement(s) de santé depuis la dernière visite?					
SOSTRIAL	SS		1	8 T30M	41	CHGTTT	Le participant a-t-il eu des changements de traitements depuis la dernière visite?					
SOSTRIAL	SS		1	8 T30M	42	HOSPIT	Depuis le dernier suivi, le participant a-t-il eu au moins une hospitalisation?					
SOSTRIAL	SS		1	8 T30M	43	DC	Le participant est-il décédé?					
SOSTRIAL	SS		1	8 T30M	44	INSTITUT	Le participant est-il en institution?					

Domaine DA (« Drug Accountability »)

STUDYID	DOMAIN	USUBJID	VISITNUM	VISIT	DASEQ	DATESTCD	DATEST	DAORRES	DASTRESC	DADTC
SOSTRIAL	DA		1	3 T3M		1 STPTTT	Le traitement IAC a-t-il été arrêté depuis la dernière visite?			
SOSTRIAL	DA		1	3 T3M		2 IACAUTRE	Si STPTTT=Oui, un autre IAC a-t-il été prescrit?			
SOSTRIAL	DA		1	3 T3M		3 CHGPOSO	La posologie du traitement a-t-elle été modifiée depuis la dernière visite?			
SOSTRIAL	DA		1	3 T3M		4 REASCHG	Si STPTTT=Oui ou CHGPOSO=Oui, pour quelle raison?			
SOSTRIAL	DA		1	3 T3M		5 REASCHGP	Si REASCHG=autre, précision			
SOSTRIAL	DA		1	4 T6M/RANDO		6 STPTTT	Le traitement IAC a-t-il été arrêté depuis la dernière visite?			
SOSTRIAL	DA		1	4 T6M/RANDO		7 IACAUTRE	Si STPTTT=Oui, un autre IAC a-t-il été prescrit?			
SOSTRIAL	DA		1	4 T6M/RANDO		8 CHGPOSO	La posologie du traitement a-t-elle été modifiée depuis la dernière visite?			
SOSTRIAL	DA		1	4 T6M/RANDO		9 REASCHG	Si STPTTT=Oui ou CHGPOSO=Oui, pour quelle raison?			
SOSTRIAL	DA		1	4 T6M/RANDO		10 REASCHGP	Si REASCHG=autre, précision			
SOSTRIAL	DA		1	4 T6M/RANDO		11 DELUNIT	Nombre d'unités d'IAC délivrées			
SOSTRIAL	DA		1	4 T6M/RANDO		12 APRUNIT	Nombre d'unités à prendre			
SOSTRIAL	DA		1	4 T6M/RANDO		13 RETUNIT	Nombre d'unités retournées			
SOSTRIAL	DA		1	4 T6M/RANDO		14 UTIUNIT	Nombre d'unités utilisées			
SOSTRIAL	DA		1	4 T6M/RANDO		15 COMPL	Compliance			
SOSTRIAL	DA		1	4 T6M/RANDO		16 COMP100	La compliance vaut-elle 100%?			
SOSTRIAL	DA		1	4 T6M/RANDO		17 COMP100N	Si COMP100=Non, raison			
SOSTRIAL	DA		1	5 T12M		18 STPTTT	Le traitement IAC a-t-il été arrêté depuis la dernière visite?			
SOSTRIAL	DA		1	5 T12M		19 IACAUTRE	Si STPTTT=Oui, un autre IAC a-t-il été prescrit?			
SOSTRIAL	DA		1	5 T12M		20 CHGPOSO	La posologie du traitement a-t-elle été modifiée depuis la dernière visite?			
SOSTRIAL	DA		1	5 T12M		21 REASCHG	Si STPTTT=Oui ou CHGPOSO=Oui, pour quelle raison?			
SOSTRIAL	DA		1	5 T12M		22 REASCHGP	Si REASCHG=autre, précision			
SOSTRIAL	DA		1	5 T12M		23 DELUNIT	Nombre d'unités d'IAC délivrées			
SOSTRIAL	DA		1	5 T12M		24 APRUNIT	Nombre d'unités à prendre			
SOSTRIAL	DA		1	5 T12M		25 RETUNIT	Nombre d'unités retournées			
SOSTRIAL	DA		1	5 T12M		26 UTIUNIT	Nombre d'unités utilisées			
SOSTRIAL	DA		1	5 T12M		27 COMPL	Compliance			
SOSTRIAL	DA		1	5 T12M		28 COMP100	La compliance vaut-elle 100%?			
SOSTRIAL	DA		1	5 T12M		29 COMP100N	Si COMP100=Non, raison			
SOSTRIAL	DA		1	6 T18M		30 STPTTT	Le traitement IAC a-t-il été arrêté depuis la dernière visite?			
SOSTRIAL	DA		1	6 T18M		31 IACAUTRE	Si STPTTT=Oui, un autre IAC a-t-il été prescrit?			
SOSTRIAL	DA		1	6 T18M		32 CHGPOSO	La posologie du traitement a-t-elle été modifiée depuis la dernière visite?			
SOSTRIAL	DA		1	6 T18M		33 REASCHG	Si STPTTT=Oui ou CHGPOSO=Oui, pour quelle raison?			
SOSTRIAL	DA		1	6 T18M		34 REASCHGP	Si REASCHG=autre, précision			
SOSTRIAL	DA		1	6 T18M		35 DELUNIT	Nombre d'unités d'IAC délivrées			
SOSTRIAL	DA		1	6 T18M		36 APRUNIT	Nombre d'unités à prendre			
SOSTRIAL	DA		1	6 T18M		37 RETUNIT	Nombre d'unités retournées			
SOSTRIAL	DA		1	6 T18M		38 UTIUNIT	Nombre d'unités utilisées			
SOSTRIAL	DA		1	6 T18M		39 COMPL	Compliance			
SOSTRIAL	DA		1	6 T18M		40 COMP100	La compliance vaut-elle 100%?			
SOSTRIAL	DA		1	6 T18M		41 COMP100N	Si COMP100=Non, raison			
SOSTRIAL	DA		1	7 T24M		42 STPTTT	Le traitement IAC a-t-il été arrêté depuis la dernière visite?			
SOSTRIAL	DA		1	7 T24M		43 IACAUTRE	Si STPTTT=Oui, un autre IAC a-t-il été prescrit?			
SOSTRIAL	DA		1	7 T24M		44 CHGPOSO	La posologie du traitement a-t-elle été modifiée depuis la dernière visite?			
SOSTRIAL	DA		1	7 T24M		45 REASCHG	Si STPTTT=Oui ou CHGPOSO=Oui, pour quelle raison?			
SOSTRIAL	DA		1	7 T24M		46 REASCHGP	Si REASCHG=autre, précision			
SOSTRIAL	DA		1	7 T24M		47 DELUNIT	Nombre d'unités d'IAC délivrées			
SOSTRIAL	DA		1	7 T24M		48 APRUNIT	Nombre d'unités à prendre			
SOSTRIAL	DA		1	7 T24M		49 RETUNIT	Nombre d'unités retournées			
SOSTRIAL	DA		1	7 T24M		50 UTIUNIT	Nombre d'unités utilisées			
SOSTRIAL	DA		1	7 T24M		51 COMPL	Compliance			
SOSTRIAL	DA		1	7 T24M		52 COMP100	La compliance vaut-elle 100%?			
SOSTRIAL	DA		1	7 T24M		53 COMP100N	Si COMP100=Non, raison			
SOSTRIAL	DA		1	8 T30M		54 STPTTT	Le traitement IAC a-t-il été arrêté depuis la dernière visite?			
SOSTRIAL	DA		1	8 T30M		55 IACAUTRE	Si STPTTT=Oui, un autre IAC a-t-il été prescrit?			
SOSTRIAL	DA		1	8 T30M		56 CHGPOSO	La posologie du traitement a-t-elle été modifiée depuis la dernière visite?			
SOSTRIAL	DA		1	8 T30M		57 REASCHG	Si STPTTT=Oui ou CHGPOSO=Oui, pour quelle raison?			
SOSTRIAL	DA		1	8 T30M		58 REASCHGP	Si REASCHG=autre, précision			
SOSTRIAL	DA		1	8 T30M		59 DELUNIT	Nombre d'unités d'IAC délivrées			
SOSTRIAL	DA		1	8 T30M		60 APRUNIT	Nombre d'unités à prendre			
SOSTRIAL	DA		1	8 T30M		61 RETUNIT	Nombre d'unités retournées			
SOSTRIAL	DA		1	8 T30M		62 UTIUNIT	Nombre d'unités utilisées			
SOSTRIAL	DA		1	8 T30M		63 COMPL	Compliance			
SOSTRIAL	DA		1	8 T30M		64 COMP100	La compliance vaut-elle 100%?			
SOSTRIAL	DA		1	8 T30M		65 COMP100N	Si COMP100=Non, raison			

Domaine AE (« Adverse Events »)

STUDYID	DOMAIN	USUBJID	AESQ	AERFID	AETERM	AEDECOD	AELC	AESER	AETOXGR	AEMAXGR	AEMAXDTC	AECN	AECNOTH	AEREL	AEOUT	AESTDTC	AEENDTC	AEONGO
SOSTRIAL	AE		1	1	123 Hospitalisation			Non		2		Aucune		Non	Aucune	06/09/2017	08/09/2017	Non

Domaine DD (« Death Details »)

STUDYID	DOMAIN	USUBJID	DDSEQ	DDTESTCD	DDTEST	DDORRES	DDSTRESC	DDDTC
SOSTRIAL	DD		1	1 INITCAUS	Cause(s) initiale(s) de décès			
SOSTRIAL	DD		1	2 OTHCAUS	Autres causes ayant contribué au décès			

ANNEXE 3 : Variables par domaine du SDTM créé pour SOStrial

Tous les domaines contiennent les variables STUDYID (identifiant de l'étude), DOMAIN (abréviation du domaine) et USUBJID (identifiant unique du participant). Les variables marquées d'une « * » ne sont pas des variables prévues par le SDTM.

Domaine	Variable SDTM	Description
SV	VISITNUM	Numéro de la visite
	VISIT	Nom de la visite
	SVSTAT	Statut de la visite
	SVDTC	Date de la visite
	SVREATEL*	Raison pour laquelle la visite a été faite par téléphone
	SVREASND	Raison pour laquelle la visite n'a pas été faite
	SVACCOMP*	Accompagnant lors de la visite
DM	BRTHDTC	Date d'anniversaire
	SEX	Sexe
	ARM	Bras de randomisation
	ARMCD	Code de randomisation
	RANDONAME*	Nom de la personne qui a effectué la randomisation
	RANDODTC	Date de la randomisation
	RFSTDTC	Date de début de suivi
	RFENDTC	Date de fin de suivi
	RFXSTDTC	Date de première prise du traitement
	RFICDTC	Date de dernière prise du traitement
	DTHFL	Indicateur de décès du participant
	DTHDTC	Date de décès du participant
SC	VISITNUM	Numéro de la visite
	VISIT	Nom de la visite
	SCSEQ	Numéro de la séquence
	SCTESTCD	Code du test
	SCTEST	Test
	SCORRES	Résultat au format numérique
	SCSTRESC	Résultat décodé le cas échéant
	EX	VISITNUM
VISIT		Nom de la visite
EXSEQ		Numéro de la séquence
EXCAT		Catégorie
EXTRT		Traitement
EXDOSE		Dose
EXDOSU		Unité de la dose
EXDOSFRM		Forme galénique
EXDOSTOT		Dose journalière
EXADJ		Raison de l'ajustement
EXSTDTC		Date de début
EXENDTC		Date de fin
CM		VISITNUM
	VISIT	Nom de la visite
	CMSEQ	Numéro de la séquence
	CMTRT	Traitement
	CMINDC	Indication
	CMPRESP	Le traitement est-il pré-spécifié dans le cahier ?

Domaine	Variable SDTM	Description
	CMOCCUR	Le participant prend-il le traitement pré-spécifié ?
	CMCLASS	Classe du traitement
	CMDOSE	Dose
	CMDOSU	Unité de la dose
	CMDOSTOT	Dose journalière
	CMDOSFRM	Forme galénique
	CMSTDTC	Début de traitement
	CMENDTC	Fin de traitement
	CMONGO	Le traitement est-il en cours à la fin de l'étude ?
TI	VISITNUM	Numéro de la visite
	VISIT	Nom de la visite
	TISEQ	Numéro de la séquence
	TITESTCD	Code du test
	TITEST	Test
	TIORRES	Résultat au format numérique
	TISTRESC	Résultat décodé le cas échéant
	TIDTC	Date
MH	MHSEQ	Numéro de la séquence
	MHTERM	Nom de l'antécédent ou pathologie
	MHDECOD	Nom standardisé
	MHPRESP	L'antécédent est-il pré-spécifié dans le cahier ?
	MHOCCUR	Le participant présente-t-il l'antécédent pré-spécifié ?
	MHSTDTC	Date de début
	MHENDTC	Date de fin
	MHENREF	La pathologie est-elle en cours à la fin de l'étude?
PE	VISITNUM	Numéro de la visite
	VISIT	Nom de la visite
	PESEQ	Numéro de la séquence
	PECAT	Catégorie
	PETEST	Code du test
	PETESTCD	Test
	PEORRES	Résultat au format numérique
	PESTRESC	Résultat décodé le cas échéant
	PEDTC	Date
	PESTAT	Statut
	PEREASND	Raison pour laquelle l'examen n'a pas été fait
VS	VISITNUM	Numéro de la visite
	VISIT	Nom de la visite
	VSSEQ	Numéro de la séquence
	VSDTC	Date
	VSTESTCD	Code du test
	VSTEST	Test
	VSORRES	Résultat au format numérique
	VSSTRESC	Résultat décodé le cas échéant
	VSSTAT	Statut
	VSREASND	Raison pour laquelle l'examen n'a pas été fait
QS	VISITNUM	Numéro de la visite
	VISIT	Nom de la visite
	QSSEQ	Numéro de la séquence

Domaine	Variable SDTM	Description
	QSDTC	Date
	QSCAT	Catégorie
	QSTESTCD	Code du test
	QSTEST	Test
	QSORRES	Résultat au format numérique
	QSSTRESC	Résultat décodé le cas échéant
	QSSTAT	Statut
	QSREASND	Raison pour laquelle le questionnaire n'a pas été fait
DA	VISITNUM	Numéro de la visite
	VISIT	Nom de la visite
	DASEQ	Numéro de la séquence
	DATESTCD	Code du test
	DATEST	Test
	DAORRES	Résultat au format numérique
	DASTRESC	Résultat décodé le cas échéant
	DADTC	Date
SS	VISITNUM	Numéro de la visite
	VISIT	Nom de la visite
	SSSEQ	Numéro de la séquence
	SSTESTCD	Code du test
	SSTEST	Test
	SSORRES	Résultat au format numérique
	SSSTRESC	Résultat décodé le cas échéant
	SSSTAT	Statut
	SSREASND	Raison pour laquelle ça n'a pas été fait
	SSDTC	Date
AE	AESEQ	Numéro de la séquence
	AEREFID	Numéro du promoteur
	AETERM	Nom de l'événement indésirable
	AEDECOD	Nom standardisé de l'événement indésirable
	AELOC	Localisation
	AESER	Evénement grave ?
	AETOXGR	Grade de toxicité
	AEMAXGR*	Grade maximal atteint
	AEMAXDTC*	Date du grade maximal atteint
	AEACN	Décision prise vis-à-vis du traitement à l'étude
	AEACNOTH	Autre action prise
	AEREL	Causalité
	AEOUT	Résultat de l'événement indésirable
	AESTDTC	Date de début
	AEENDTC	Date de fin
	AEONGO	L'événement est-il en cours à la fin de l'étude ?
DD	DDSEQ	Numéro de séquence
	DDTESTCD	Code du test
	DDTEST	Test
	DDORRES	Résultat numérique
	DDSTRESC	Résultat décodé le cas échéant
	DDDTC	Date

ANNEXE 4 : Paramétrage des tables sous CSExport

Nom de la table :	VS	Nombre de variables :	17	
Date de création :	03/04/2019 15:36:03	Nombre d'occurrences :	56	
Description :	Examen clinique			
---- Liste des variables par ordre alphabétique ----				
#	Variable	Long.	Format	Intitulé
9	ARM	254	254.	Groupe de randomisation
10	ARMCD	254	254.	Code de randomisation
4	CENTERID	2	8.	Numéro du centre
2	DOMAIN	2	10.	Domaine
7	SEQNUM	2	2.	Numéro de séquence
1	STUDYID	8	75.	Nom de l'étude
3	USUBJID	254	254.	Code Patient
5	VISIT	254	254.	Nom de la visite
8	VISITDY	10	10.	Date de la visite
6	VISITNUM	1	2.	Numéro de la visite
17	VSFMT	7	10.	Format de VSORRES
13	VSORRES	8	8.	Résultat numérique
15	VSORRESU	4	10.	Unité
14	VSSTRESC	8	200.	Résultat caractères
11	VSTEST	44	200.	Libellé du test
12	VSTESTCD	6	10.	Code du test
16	VSTEXT	39	200.	Précision

Nom de la table :	QS	Nombre de variables :	18	
Date de création :	03/04/2019 15:34:10	Nombre d'occurrences :	216	
Description :	Questionnaires			
---- Liste des variables par ordre alphabétique ----				
#	Variable	Long.	Format	Intitulé
9	ARM	254	254.	Groupe de randomisation
10	ARMCD	254	254.	Code de randomisation
4	CENTERID	2	8.	Numéro du centre
2	DOMAIN	2	10.	Domaine
11	QSCAT	200.		Catégorie du questionnaire
18	QSFMT	7	7.	Format du résultat
15	QSORRES	3	8.	Résultat numérique
12	QSSCAT	38	200.	Sous catégorie du questionnaire
16	QSSTRESC	147	147.	Résultat caractère
17	QSSTRESN	3	8.	Résultat du champ calculé
13	QSTEST	147	200.	Libellé du test
14	QSTESTCD	7	10.	Code du test
7	SEQNUM	3	8.	Numéro de séquence
1	STUDYID	10	75.	Nom de l'étude
3	USUBJID	254	254.	Identifiant du sujet
5	VISIT	254	254.	Libellé de la visite
8	VISITDY	10	10.	Date de la visite
6	VISITNUM	2	2.	Numéro de visite

Nom de la table :	MH	Nombre de variables :	11	
Date de création :	05/04/2019 14:02:59	Nombre d'occurrences :	25	
Description :	Antécédents			
---- Liste des variables par ordre alphabétique ----				
#	Variable	Long.	Format	Intitulé
2	DOMAIN	2	10.	Domaine
10	MHENDTC	10	10.	Date de fin
11	MHONGO	3	F01001_	En cours à la fin de l'étude
7	MHSEQ	2	2.	Numéro de la séquence
9	MHSTDTC	10	10.	Date de début
8	MHTERM	12	250.	Nom de l'antécédent
1	STUDYID	10	75.	Nom de l'étude
3	USUBJID	254	254.	Identifiant du sujet
4	VISIT	254	254.	Visite
6	VISITDY	10	10.	Date de la visite
5	VISITNUM	2	1.	Numéro de la visite

Nom de la table : SC **Nombre de variables :** 17
Date de création : 03/04/2019 15:35:01 **Nombre d'occurrences :** 43
Description : Caractéristiques du sujet

---- Liste des variables par ordre alphabétique ----

#	Variable	Long.	Format	Intitulé
9	ARM		75.	Groupe de randomisation
10	ARMCD		10.	Code de randomisation
4	CENTERID	2	8.	Numéro du centre
2	DOMAIN	2	10.	Domaine
11	SCCAT	28	75.	Catégorie
17	SCFMT	7	7.	Format du résultat
14	SCORRES	10	Best12.	Résultat
15	SCSTRESC	254	3900.	Résultat caractères
12	SCTEST	91	200.	Libellé du test
13	SCTESTCD	4	10.	Code du test
16	SCTEXT	0	3900.	Précision
7	SEQNUM	2	2.	Numéro de séquence
1	STUDYID	10	75.	Nom de l'étude
3	USUBJID	254	254.	Code Patient
5	VISIT	254	254.	Nom de la visite
8	VISITDY	10	10.	Date de la visite
6	VISITNUM	1	1.	Numéro de la visite

Nom de la table : IE **Nombre de variables :** 16
Date de création : 03/04/2019 15:35:31 **Nombre d'occurrences :** 22
Description : Critères d'inclusion/exclusion

---- Liste des variables par ordre alphabétique ----

#	Variable	Long.	Format	Intitulé
5	ARM		10.	Bras de randomisation
6	ARMCD		10.	Code de randomisation
4	CENTERID	2	8.	Numéro du centre
2	DOMAIN	2	10.	Domaine
11	IECAT	28	75.	Catégorie
16	IEFMT	6	10.	Format du résultat
14	IEORRES	1	8.	Résultat numérique
15	IESTRESC	3	10.	Résultat caractères
13	IETEST	166	200.	Critère
12	IETESTCD	4	4.	Code du critère
9	SEQNUM	2	2.	Numéro de séquence
1	STUDYID	254	254.	Nom de l'étude
3	USUBJID	254	254.	Code Patient
7	VISIT	254	254.	Nom de la visite
10	VISITDY	10	10.	Date de la visite
8	VISITNUM	1	1.	Numéro de la visite

Nom de la table : CM **Nombre de variables :** 19
Date de création : 07/06/2019 15:18:52 **Nombre d'occurrences :** 92
Description : Traitements concomitants

---- Liste des variables par ordre alphabétique ----

#	Variable	Long.	Format	Intitulé
9	CMCLASS	16	F99035_.	Catégorie du traitement
16	CMDOSE	6	8.	Dose journalière du médicament
17	CMDOSEU	15	15.	Unité de la dose de médicament
18	CMDOSEUO	5	15.	Précision unité
14	CMENDTC	10	10.	Date de fin
15	CMFORM	8	F13011_.	Forme galénique du traitement
10	CMIND		75.	Indication du traitement
12	CMOCCUR	3	OUIINON.	Le sujet a-t-il pris le médicament spécifié?
19	CMONGO	3	F01001_.	En cours à la fin de l'étude
11	CMPRESP	3	10.	Médicament pré-spécifié sur l'eCRF?
7	CMSEQ	2	2.	Numéro de la séquence
13	CMSTDTC	10	10.	Date de début
8	CMTERM	12	250.	Nom du traitement
2	DOMAIN	2	10.	Domaine
1	STUDYID	10	75.	Nom de l'étude
3	USUBJID	254	254.	Identifiant du sujet
4	VISIT	254	254.	Visite
6	VISITDY	10	10.	Date de la visite
5	VISITNUM	2	1.	Numéro de la visite

Nom de la table : DS
Date de création : 03/04/2019 15:36:33
Description : Suivi du sujet

Nombre de variables : 17
Nombre d'occurrences : 39

---- Liste des variables par ordre alphabétique ----

#	Variable	Long.	Format	Intitulé
4	ARM		10.	Bras de randomisation
5	ARMCD		10.	Code de randomisation
3	CENTERID	2	8.	Numéro du centre
2	DOMAIN	2	10.	Domaine
11	DSCAT	13	75.	Catégorie
17	DSFMT	7	10.	Format du résultat
14	DSORRES	10	Best12.	Résultat numérique
15	DSSTRESC	254	254.	Résultat caractères
12	DSTEST	166	200.	Libellé du test
13	DSTESTCD	4	10.	Code du test
16	DSTEXT	30	200.	Précision
10	SEQNUM	2	2.	Numéro de séquence
1	STUDYID	254	254.	Nom de l'étude
9	USUBJID	254	254.	Code Patient
6	VISIT	254	254.	Nom de la visite
8	VISITDY		10.	Date de la visite
7	VISITNUM	1	1.	Num de la visite

Nom de la table : AE
Date de création : 25/02/2019 14:41:45
Description : Effets indésirables

Nombre de variables : 17
Nombre d'occurrences : 40

---- Liste des variables par ordre alphabétique ----

#	Variable	Long.	Format	Intitulé
15	AEACN	28	28.	Action sur le traitement
16	AEACNOTH	22	250.	Autre(s) action(s)
8	AEENDTC	10	10.	Date de fin
13	AEMAXDTC	10	10.	Date du grade maximal
9	AEONGO	3	3.	En cours à la fin de l'étude
17	AEOUT	25	25.	Evolution
5	AEREFID	10	10.	Si grave, numéro EIG promoteur
14	AEREL	13	13.	Imputabilité à l'étude
4	AESEQ	3	8.	Numéro de séquence
10	AESER	9	9.	Gravité de l'EI
7	AESTDTC	10	10.	Date de début
6	AETERM	21	250.	Nom de l'événement indésirable
12	AETGRMAX	1	8.	Grade maximal atteint
11	AETOXGR	1	8.	Grade de l'événement
2	DOMAIN	2	2.	Domaine
1	STUDYID	8	8.	Nom de l'étude
3	USUBJID	254	254.	Identifiant du sujet

Nom de la table : EX
Date de création : 05/04/2019 14:03:04
Description : Traitement à l'étude

Nombre de variables : 15
Nombre d'occurrences : 7

---- Liste des variables par ordre alphabétique ----

#	Variable	Long.	Format	Intitulé
2	DOMAIN	2	10.	Domaine
15	EXADJ	21	F06057_.	Raison de l'ajustement
9	EXCLASS	16	250.	Catégorie du traitement
13	EXDOSETOT	6	8.	Dose journalière du médicament
14	EXDOSEU	15	15.	Unité de la dose de médicament
11	EXENDTC	10	10.	Date de fin
12	EXFORM	8	F13011_.	Forme galénique du traitement
7	EXSEQ	2	2.	Numéro de la séquence
10	EXSTDTC	10	10.	Date de début
8	EXTRT	12	10.	Nom du traitement
1	STUDYID	10	75.	Nom de l'étude
3	USUBJID	254	254.	Identifiant du sujet
4	VISIT	254	254.	Visite
6	VISITDY	10	10.	Date de la visite
5	VISITNUM	2	1.	Numéro de la visite

ANNEXE 5 : Cahier des charges pour des tables fournies pour l'analyse

	Entité d'application : USMR Emetteur : USMR	EN-USM-809
	DOCUMENT D'ENREGISTREMENT	Ind : 0.2 Page : 1/4
Cahier des charges		

	Entité d'application : USMR Emetteur : USMR	EN-USM-809
	DOCUMENT D'ENREGISTREMENT	Ind : 0.2 Page : 2/4
Cahier des charges		

Objet :

Définition des tables à fournir pour l'analyse à l'USMR dans le cadre d'une étude en CDISC.

Domaine d'application :

Gestion de données, analyses statistiques.

Cahier des charges :**I. Contexte**

Le CDISC (*Clinical Data Interchange Standards Consortium*) est un consortium international à but non lucratif qui vise à promouvoir la standardisation des formats de recueil, d'échange, de soumission et d'archivage de données dans la recherche clinique. Ces formats sont utilisés comme référence dans le cadre d'échange de données (soumission à la FDA par exemple).

L'USMR gère une multitude de bases de données portant sur des domaines médicaux différents. Des efforts ont déjà été mis en place pour standardiser l'organisation des informations et des réflexions ont été menées sur l'application d'un standard international facilitant la transmission de données ou l'application de programmes standards (projet ADaM notamment).

Un projet s'inscrivant dans le processus « recherche et développement » avait par ailleurs été mis en place par Olivier QUINTIN afin d'effectuer une étude (SOSrial) en CDISC. L'étude ayant été stoppée, le projet n'avait pas pu aboutir. Pauline GAFFEZ l'a repris comme support pour son sujet de recherche de master.

II. Objectif

L'objectif de ce cahier des charges est d'exprimer les besoins des statisticiens concernant les tables qui seraient fournies après transformation des tables CDISC pour l'analyse statistique. Les demandes exprimées seront examinées pour voir si elles sont réalisables ou non, selon leur degré de complexité.

Le but est de développer un programme incluant les demandes des statisticiens, dans la limite de leur faisabilité.

III. Description des pratiques actuelles à l'USMR

Les données aujourd'hui fournies pour l'analyse sont stockées sous forme de tables. Les tables et les variables les plus communes sont définies dans un document-type qui sert de support à la création de la base de données, il s'agit du document d'organisation des bases (DOB).

Les tables et les variables les plus communes sont définies dans un document-type qui sert de support à la création de la base de données, il s'agit du document d'organisation des bases (DOB).

Une convention de nommage a également été mise en place pour les variables les plus courantes telles que le poids, la taille, les critères d'inclusion/non inclusion, etc. Cette convention permet une certaine standardisation des noms de variables (par exemple « PDS » pour le poids).

Des programmes standards d'analyses statistiques sont paramétrés et tournent sur les tables fournies aux statisticiens afin de réaliser les analyses. Ces tables fournies doivent respecter certaines contraintes pour que les programmes standards fonctionnent correctement (toutes les variables en numérique avec un format SAS associé). Ces programmes ont été développés sous SAS et ont en partie été qualifiés dans une démarche qualité, afin d'assurer la fiabilité des résultats obtenus.

IV. Expression des besoins

La priorité est donnée pour avoir toutes les données en « 1 ligne par sujet » dans les tables, comme c'est le cas actuellement. Il faudra donc transposer les tables CDISC tout en gardant l'information de la visite d'où provient la donnée, par exemple grâce à l'ajout d'un préfixe. Ensuite, si cela est réalisable, les statisticiens pourraient avoir besoin des données en « 1 ligne par visite » dans une même table.

	Entité d'application : USMR Emetteur : USMR	EN-USM-809
	DOCUMENT D'ENREGISTREMENT	Ind : 0.2 Page : 3/4
Cahier des charges		

Il faut récupérer les observations au format numérique avec, lorsque cela est pertinent, le format associé pour pouvoir décoder les variables sous SAS. Cela concerne donc toutes les variables de classe, i.e. toutes celles qui ne seront pas analysables de manière purement quantitative (ex : poids, taille, etc.). Des essais seront menés pour voir quelle solution est la plus efficace afin de récupérer les formats des variables avec les tables au format CDISC. Eventuellement, des formats pourront être créés même pour les variables numériques, car en pratique les statisticiens peuvent être amenés à le faire.

Le label des variables doit être gardé car ce sont eux qui apparaissent dans les rapports d'analyse. Les noms donnés aux variables sont exploités dans les programmes et ne sont pas affichés dans les rapports.

Il apparaît nécessaire de définir des conventions de nommage, tant pour les noms des tables, que des variables, que des noms des visites. A cette fin, la convention de nommage existante doit être mise à jour et complétée afin de standardiser les pratiques. Cela pourra faire l'objet de discussions avec les statisticiens.

V. Contraintes

- La programmation doit être effectuée sous SAS.
- Le programme doit être étude-indépendant, il doit fonctionner quelles que soient les tables exportées.
- Le paramétrage du programme doit être relativement simple.
- Il ne faut pas que le paramétrage de l'eCRF et de l'export devienne trop complexe.

VI. Responsabilité

Rémi SITTA définira les besoins exprimés par les statisticiens tandis que Pauline GAFFEZ s'occupera de l'étude de la faisabilité des demandes exprimées et de la réalisation du programme SAS.

ANNEXE 6 : Code du programme Croche-patte

Fichier « v0.1_Croche-patte_31052019.sas »

```

/*****
/***** PROGRAMME Croche-patte v0.1 *****/
/***** 31/05/2019 PAULINE GAFFEZ *****/
/*****/

* Croche-patte est un programme développé en interne à l'USMR qui permet de
transposer des tables initialement au format CDSIC pour obtenir des tables
dans un format horizontal. Il est initialement développé pour tourner sur
des tables issues de CSExport.;

* Pour l'utiliser, il faut au préalable définir les paramètres ci-dessous :
- chemins d'import et export des tables
- tables que l'on souhaite transposer
- variables communes à toutes les tables et que l'on souhaite garder
- variables CDISC que l'on souhaite transposer pour chaque table
- suffixe qui sera ajouté au nom des variables finales pour
distinguer les différentes variables CDISC transposées (ex: POIDS_R pour le
résultat original, POIDS_P pour une précision, etc.)
- précision qui apparaîtra dans le label original (ex: "Examen
clinique (résultat)", "Examen clinique (précision)", etc.)
- dérivée et module(s) à lancer
;

* Pour plus d'explications, reportez-vous à la notice d'utilisation du
programme.;

*****

* Chemin où se trouvent les tables à transposer et les formats;
%let strChemin=R:\EQUIPES\USMR\DATA MANAGEMENT\Stagiaire\Pauline
GAFFEZ\CDISC_SOS_TRIAL\Programmation\Croche-patte - v0.1 - 31052019\tables;

* Nom du fichier où se trouvent les formats;
%let formats=RES_ORACLE_EXCEL_20190307_135941.xls;

* Tables CDISC à transposer (le nom est celui donné dans CSExport);
%let tab1=VS;
%let tab2=QS;
%let tab3=DS;

* Variables communes à toutes les tables à garder;
%let var_com=STUDYID USUBJID ARM ARMCD /*VISIT VISITNUM VISITDY*/;

* Variables CDISC, suffixes et labels à utiliser pour la table VS ;
%let VS_var1=VSORRES;
%let VS_var2=VSTEXT;
%let VS_var3=VSORRESU;
%let VS_s1=_R;
%let VS_s2=_P;
%let VS_s3=_U;
%let VS_lab1=résultat ;
%let VS_lab2=précision ;
%let VS_lab3=unité ;

* Variables CDISC, suffixes et labels à utiliser pour la table QS ;
%let QS_var1=QSORRES;

```

```

%let QS_var2=QSSTRESN;
%let QS_s1=_R;
%let QS_s2=_C;
%let QS_lab1=résultat ;
%let QS_lab2=champ calculé ;

* Variables CDISC, suffixes et labels à utiliser pour la table DS ;
%let DS_var1=DSORRES;
%let DS_var2=DSTEXT;
%let DS_s1=_R;
%let DS_s2=_P;
%let DS_lab1=résultat ;
%let DS_lab2=précision ;

* Dérivée et modules optionnels à lancer ;
* /\ Le recours au module B permet de supprimer toutes les colonnes sans
aucune donnée(colonnes construites par le programme). Il peut donc amener à
supprimer des colonnes pour lesquelles aucune donnée n'est disponible.;

%let module_A=0; * 1 ligne par patient par visite (1=oui 0=non);
%let module_B=1; * Suppression des colonnes vides (1=oui 0=non);
%let derivee=1; * 1=export brut 2=formats 3=libellés 4=formats et libellés;

*****
----- NE PAS MODIFIER -----

* Fichiers du programme ;
%let strCheminP=R:\EQUIPES\USMR\DATA MANAGEMENT\Stagiaire\Pauline
GAFFEZ\CDISC_SOS_TRIAL\Programmation\Croche-patte - v0.1 -
31052019\Fichiers Croche-patte v0.1 31052019;

* Vérification MD5 et appel du programme ;
%let temps_debut = %sysfunc(time());
%let empreinte_ref=6DAFECF3F0E6E125D09C70CA37BD3913 ;
options noxwait; X "R: & ""&strCheminP.\Batch_cleMd5-v01-20170303.bat""";

data empreinte; length empreinte $32; infile "&strCheminP.\md5.txt" ; input
empreinte 1-32; run;
proc sql noprint;
    select * into: empreinte_calc from empreinte;
    drop table empreinte;
quit;

options noxwait; X "del ""&strCheminP.\md5.txt""";

%macro appel;

    %if "&empreinte_calc"="&empreinte_ref" %then
        %do; %include "&strCheminP.\v0.1_import_31052019.sas";
        %include "&strCheminP.\v0.1_main_31052019.sas";
        %include "&strCheminP.\v0.1_rapport_31052019.sas"; %end;
    %else
        %do; %put ERROR: échec de l authentification. La version
utilisée n est pas celle qui a été qualifiée.;
        %abort; %end;

%mend appel;
%appel

```

Fichier « v0.1_Import_31052019.sas »

```
* Récupération du nombre de tables paramétrées;

proc sql noprint;
    select count(name) into: nb_tab
        from sashelp.vmacro
        where name like "TAB%";
quit;

* Import des tables à transposer;

%macro import;
    filename dossier "&strChemin.";
    %do idx_tab=1 %to &nb_tab.;
        %include dossier(&&tab&idx_tab...sas);
    %end;
    proc copy in=CDISC out=work;
    run;
%mend import;
%import
```

Fichier « v0.1_main_31052019.sas »

```
%macro main;

%if (&derivee. ne 1 AND &derivee. ne 2 AND &derivee. ne 3 AND &derivee. ne
4)
    OR (&module_A. ne 0 AND &module_A. ne 1) OR (&module_B. ne 0 AND
&module_B. ne 1)
%then
    %do; %put ERROR: Il y a une erreur dans le paramétrage de la dérivée
ou des modules. ; %abort; %end;
%else
    %do;
        * Appel du moteur;
        %include "&strCheminP.\v0.1_moteur_31052019.sas";
        * Appel du module A (optionnel);
        %if &module_A.=1
            %then %do; %include
"&strCheminP.\v0.1_A_tab_semihorizontales_31052019.sas"; %end;
        * Appel de la dérivée;
        %if &derivee.=1
            %then %do; %include
"&strCheminP.\v0.1_1_export_brut_31052019.sas"; %end;
        %if &derivee.=2
            %then %do; %include
"&strCheminP.\v0.1_2_formats_31052019.sas"; %end;
        %if &derivee.=3
            %then %do; %include
"&strCheminP.\v0.1_3_libelles_31052019.sas"; %end;
        %if &derivee.=4
            %then %do; %include
"&strCheminP.\v0.1_4_formats_libelles_31052019.sas"; %end;
        * Appel du module B (optionnel);
        %if &module_B.=1
            %then %do; %include
"&strCheminP.\v0.1_B_supp_col_31052019.sas"; %end;
    %end;
```

```
%mend main;
%main
```

Fichier « v0.1_moteur_31052019.sas »

```
filename fLOG "&strChemin.\_log";
proc printto log=fLOG new; run;
options formchar="|----|+|---+|=|-\<>*" nodate;
title j=1 "Programme Croche-patte v0.1 31052019";
ods pdf
file="&strChemin.\rapport_derivee&derivee._%sysfunc(today(),ddmmyyn.).pdf";

* Récupération du nombre de tables paramétrées;

proc sql noprint;
  select count(name) into: nb_tab
  from sashelp.vmacro
  where name like "TAB%";
quit;

* Transposition des tables;

%macro transposition;

* Définition des macro-variables utilisées par le programme (visites) ;

%do idx_tab=1 %to &nb_tab;
  proc sql noprint ;
    create table LISTE_VISITE_&&tab&idx_tab. as
    select distinct VISITNUM, VISIT
    from &&tab&idx_tab.
    order by VISITNUM;
    select count(distinct VISIT) into:&&tab&idx_tab._nb_vis
    from &&tab&idx_tab. ;
  quit;
  data _null_ ;
    set liste_visite_&&tab&idx_tab.;
    cnt+1;
    call symput
(cats("&&tab&idx_tab..vis",put(cnt,best.)),compress(cats(VISIT),'-', 's'));
  run;
%end;

* Eclatement de la table par visites ;

%do idx_tab=1 %to &nb_tab;
  %do idx_vis=1 %to &&&&&tab&idx_tab._nb_vis. ;
    data &&tab&idx_tab._&&&&&tab&idx_tab..vis&idx_vis.;
      set &&tab&idx_tab.;
      if compress(visit,'-
', 's')="&&&&&tab&idx_tab..vis&idx_vis."
      then output
&&tab&idx_tab._&&&&&tab&idx_tab..vis&idx_vis.;
    run;
  %end;
%end;

* Récupération du nombre de variables définies pour la transposition;
```

```

%do idx_tab=1 %to &nb_tab;
  proc sql noprint;
    select count(name) into: &&tab&idx_tab.._nb_var
    from sashelp.vmacro
    where name like ("&&tab&idx_tab.._VAR%");
  quit;
%end;

* Transposition de toutes les tables;

%do idx_tab=1 %to &nb_tab.;
  %do idx_var=1 %to &&&&&tab&idx_tab.._nb_var.;
    %do idx_vis=1 %to &&&&&tab&idx_tab.._nb_vis.;
      proc transpose
data=&&tab&idx_tab.._&&&&&tab&idx_tab..vis&idx_vis.

      out=&&&&&tab&idx_tab.._var&idx_var.._&&&&&tab&idx_tab..vis&idx_vis.
(drop=_label__name_)
      prefix=&&&&&tab&idx_tab..vis&idx_vis.._
suffix=&&&&&tab&idx_tab.._s&idx_var. ;
      id &&tab&idx_tab..TESTCD;
      idlabel &&tab&idx_tab..TEST;
      var &&&&&tab&idx_tab.._var&idx_var.;
      by &var_com.;
    run;
  %end;
%end;
%end;

* Merge des tables récupérées ;

%do idx_tab=1 %to &nb_tab;
  * Regroupement des différentes variables ;
  %do idx_vis=1 %to &&&&&tab&idx_tab.._nb_vis.;
    data T_&&tab&idx_tab.._&&&&&tab&idx_tab..vis&idx_vis.;
      merge
        %do idx_var=1 %to &&&&&tab&idx_tab.._nb_var.;
          &&&&&tab&idx_tab.._var&idx_var.._&&&&&tab&idx_tab..vis&idx_vis.
        %end;
        ;
      by USUBJID;
    run;
  %end;

  * Puis regroupement des différentes visites ;
  data T_&&tab&idx_tab.;
    merge
      %do idx_vis=1 %to &&&&&tab&idx_tab.._nb_vis.;
        T_&&tab&idx_tab.._&&&&&tab&idx_tab..vis&idx_vis.
      %end;
    ;
    by USUBJID;
  run;
%end;

%mend transposition;
%transposition
Fichier « v0.1_A_tab_semi-horizontales_31052019.sas »

%macro tab_semi_horiz;

```

* Suppression des tables T_XX créées par le moteur ;

```
proc datasets lib=work nolist;
  delete
  %do idx_tab=1 %to &nb_tab.;
    T_&&tab&idx_tab.
  %end;
;
quit;
```

* Définition des macro-variables utilisées par le programme (visites) ;

```
%do idx_tab=1 %to &nb_tab;
  proc sql noprint ;
    create table LISTE_VISITE_&&tab&idx_tab. as
      select distinct VISITNUM, VISIT
      from &&tab&idx_tab.
      order by VISITNUM;
    select count(distinct VISIT) into:&&tab&idx_tab.._nb_vis
    from &&tab&idx_tab. ;
  quit;
  data _null_ ;
    set liste_visite_&&tab&idx_tab.;
    cnt+1;
    call symput
(cats("&&tab&idx_tab..vis",put(cnt,best.)),compress(cats(VISIT),'-', 's'));
  run;
%end;
```

* Suppression des préfixes avec les noms de visite dans les noms de variables (avant merge des visites);

```
%do idx_tab=1 %to &nb_tab.;
  %do idx_vis=1 %to &&&&&tab&idx_tab.._nb_vis.;
    proc contents
data=T_&&tab&idx_tab.._&&&&&tab&idx_tab..vis&idx_vis. noprint
out=_contents ;
      run ;
      data _null_ ;
        set _contents end=LAST;
        where name like
cats("&&&&&tab&idx_tab..vis&idx_vis.", "%") /*and name contains "_"*/;
        if _n_=1 then call execute ('proc datasets nolist;
modify T_&&tab&idx_tab.._&&&&&tab&idx_tab..vis&idx_vis. ; rename ');
        call execute ( name !! "=" !!
cats(tranwrd(name, "&&&&&tab&idx_tab..vis&idx_vis.._", " ")) );
        if last then call execute (';quit;');
      run ;
    %end;
  %end;
```

* Fusion des différentes visites à la suite ;

```
%do idx_tab=1 %to &nb_tab.;
  data T_&&tab&idx_tab.;
  set
  %do idx_vis=1 %to &&&&&tab&idx_tab.._nb_vis.;
    T_&&tab&idx_tab.._&&&&&tab&idx_tab..vis&idx_vis.
  %end;
;
;
```

```

run;
proc sort data=T_&&tab&idx_tab.;
by USUBJID VISITNUM;
run;
%end;

%mend tab_semi_horiz;
%tab_semi_horiz

```

Fichier « v0.1_4_formats_libellés_31052019.sas »

```

%macro formats_libelles;

* Import des formats;

proc import datafile="&strChemin.&formats."
out=formats_n (rename=(B=FMTNAME C=START D=LABEL)) dbms=xls
replace;
getnames=no;
datarow=6;
run;
data formats_c;
set formats_n;
where substr(FMTNAME,1,1)^="$";
FMTNAME=cats("$",FMTNAME);
run;
data formats;
set formats_n formats_c;
run;
proc format cntlin=formats; run;

* Définition des macro-variables (variables et formats associés) utilisées
par le programme ;

%do idx_tab=1 %to &nb_tab.;
proc sql noprint ;
create table &&tab&idx_tab.._FMT as
select distinct &&tab&idx_tab..TESTCD,
&&tab&idx_tab.._FMT
from &&tab&idx_tab.
where &&tab&idx_tab.._FMT is not null;
select count(distinct &&tab&idx_tab..TESTCD) into:
&&tab&idx_tab.._nb_obs_f
from &&tab&idx_tab.._fmt;
quit ;
data _null_ ;
set &&tab&idx_tab.._FMT;
cnt+1;
call symput
(cats("&&tab&idx_tab..obs_f",put(cnt,best.)),cats(&&tab&idx_tab..TESTCD));
/* observations avec un format */
call symput
(cats("&&tab&idx_tab..fmt",put(cnt,best.)),cats(&&tab&idx_tab.._FMT)); /*
formats */
run;
%end;

* Attribution des formats pour les variables en ayant un ;

```



```

%do idx_tab=1 %to &nb_tab.;
  proc contents data=T_&&tab&idx_tab noprint out=_contents ; run
;
  data _null_ ;
  %do idx_fmt=1 %to &&&&&tab&idx_tab.._nb_obs_f.;
    set _contents end=LAST;
    where name like
cats("%", "&&&&&tab&idx_tab..obs_f&idx_fmt.&&&&&tab&idx_tab..s1" ) ;
    if _n_=1 then call execute ('proc datasets lib=WORK
nolist; modify T_&&tab&idx_tab ; format ' ) ;
    if type=1 then call execute ( name !! " " !!
"&&&&&tab&idx_tab..fmt&idx_fmt..." ) ; * Formats numériques;
    else if type=2 then call execute ( name !! " "
!! cats("$", "&&&&&tab&idx_tab..fmt&idx_fmt..." ) ; * Formats caractères;
    if last then call execute (';quit;');
  %end;
run ;
%end;

* Récupération du nombre de variables définies pour la transposition;

%do idx_tab=1 %to &nb_tab;
  proc sql noprint;
    select count(name) into: &&tab&idx_tab.._nb_var
    from sashelp.vmacro
    where name like ("&&tab&idx_tab.._VAR%");
  quit;
%end;

* Attribution des libellés ;

%do idx_tab=1 %to &nb_tab.;
  proc contents data=T_&&tab&idx_tab noprint out=_contents ;
  run ;
  %do idx_var=1 %to &&&&&tab&idx_tab.._nb_var.;
  data _null_ ;
    set _contents end=LAST;
    where name like
cats("%", "&&&&&tab&idx_tab..s&idx_var.") and name contains "_";
    if _n_=1 then call execute ('proc datasets nolist; modify
T_&&tab&idx_tab ; label ' ) ;
    call execute ( name !! "=" !!
reverse(scan(reverse(name),3,'_')) !! " " !! catx(" ",translate(label,"
","'");"),("&&&&&tab&idx_tab..lab&idx_var.") ) ;
    if last then call execute (';quit;');
  %end;
run ;
%end;

* Export des tables transposées ;

libname transpo "&strChemin.";
options FMTSEARCH=(work transpo);

proc copy in=work out=transpo;
  %do idx_tab=1 %to &nb_tab.;
    select T_&&tab&idx_tab. / memtype=data;
  %end;
  select formats / memtype=catalog;
run;

```

```

    %put _global_;
    %put _local_;

%mend formats_libelles;
%formats_libelles

* Titre du rapport pdf;

ods escapechar='^';
title 'Programme Croche-patte v0.1 31052019'
'^2n _Rapport dérivée n°4_'
'^2n Application des formats, des libellés et export des tables
transposées';

```

Fichier « v0.1_B_supp_col_31052019.sas »

```

* Suppression des colonnes sans aucune donnée;

%macro colonnes;

    %do idx_tab=1 %to &nb_tab.;
        ods select none;
        ods output nlevels=temp&idx_tab.;
        proc freq data=transpo.T_&&tab&idx_tab. nlevels;
            tables _all_;
        run;
        proc sql;
            select TableVar into : drop&idx_tab. separated by ','
            from temp&idx_tab.
            where NNonMissLevels=0;
            alter table transpo.T_&&tab&idx_tab.
            drop &&drop&idx_tab.;
        quit;
    %end;

ods select close;
ods output close;

%mend colonnes;
%colonnes

```

Fichier « v0.1_rapport_31052019.sas »

```

* Génération du rapport pdf;
proc sql noprint;
    select count(name) into: nb_tab
    from sashelp.vmacro
    where name like "TAB%";
quit;

%let temps_fin = %sysfunc(time());
%let duree = %sysevalf(&temps_fin.-&temps_debut., floor);

proc printto; run;

proc sql noprint outobs=&nb_tab.;
    select cats("'",value,"'") into:tables_traitees separated by ", "

```

```

    from sashelp.Vmacro
    where name like "TAB%";
proc sql noprint;
    select value into:variables_traitees separated by ", "
    from sashelp.Vmacro
    where name like cats("%","_VAR%")
    and scan(name,1,'_') IN (&tables_traitees.);
quit;

data _null_;
    file print;
    put " ";
    put "Date et heure d'execution: %sysfunc(today(),FRADFWDX.)
%sysfunc(time(),HHMM)";
    put "Dossier: &strChemin";
    put "Bibliothèque d'export: transpo";
    put "Durée d'exécution: inférieure à &duree secondes";
    put "Tables traitées: &tables_traitees ";
    put "Variables communes à toutes les tables: &var_com";
    put "Les variables CDISC traitées sont les suivantes:";
    put "&variables_traitees";
    put "Module A utilisé: &module_A. (0=non 1=oui)";
    put "Module B utilisé: &module_B. (0=non 1=oui)";

data work._log;
    infile fLOG;
    input @1 mLOG $100.;
    if substr(mLog,1,6) = "ERROR:";
run;
proc print data=work._log; run;

%macro DD;
    %do idx_tab=1 %to &nb_tab.;
    proc sql ;
        title "Dictionnaire de données de la table T_&&tab&idx_tab.";
        select memname, name, type, length, varnum, label, format,
informat
            from sashelp.Vcolumn
            where libname="TRANSPO" and
            memname="T_&&tab&idx_tab.";
    quit;
    %end;
%mend DD;
%DD

proc document name=mydoc(write); import textfile=fLOG to logfile; run;
replay; quit;
options noxwait; X "del ""&strChemin.\_log""";
ods pdf close;

* Nettoyage de la session ;

*** work;
proc datasets lib=work kill noprint; run; quit;

*** Variables système;
%let syscc = 0;
%let sysrc = 0;
%let syslibrc = 0;
%let sysfilrc = 0;
%let syslckrc = 0;

```

```
%let syslast = ;
options obs = max;

*** Macros-variables;
proc sql noprint;
    select name into: vars separated by ' '
    from dictionary.macros
    where scope='GLOBAL'
    and not name contains 'SYS_SQL_IP_';
quit;

%symdel &vars.;


*** Filename;
proc sql noprint;
    select unique fileref into :myfiles separated by '; filename '
    from dictionary.extfiles
    where substr(fileref,1,1) ^= '#';
quit;
filename &myfiles ;

*** Macros;
proc catalog
    catalog = work.sasmacr force;
    save appel (et=macro);
    save main (et=macro);
run;
quit;

*** Titre;
title;

*** ODS;
ods _all_ close;
```

ANNEXE 7 : Notice d'utilisation du programme Croche-patte

	Entité d'application : USMR Emetteur : USMR	IN-USM-
	INSTRUCTION	Ind : 0.1 Page : 1/4
Notice d'utilisation du programme Croche-patte v0.1		

	Entité d'application : USMR Emetteur : USMR	IN-USM-
	INSTRUCTION	Ind : 0.1 Page : 2/4
Notice d'utilisation du programme Croche-patte v0.1		

Objet

Notice d'utilisation du programme Croche-patte v0.1 31052019.

Instruction

I. BUT

Le programme Croche-patte permet de transposer des tables CDISC dans un format complètement horizontal. Il est destiné à traiter des tables obtenues grâce à CSExport dans le cadre d'une étude menée dès le départ en CDISC mais peut potentiellement être utilisé sur d'autres tables.

II. PREALABLE

- Identifier la/les dérivée(s) et le(s) module(s) dont on a besoin



- Exporter les tables à l'aide du module CSExport

Dans le module CSExport, un nom de bibliothèque ainsi qu'un chemin où la créer sont à renseigner avant d'exécuter l'export. Le nom de la bibliothèque pour l'utilisation de Croche-patte est **CDISC** et le chemin est au choix **mais doit correspondre au dossier qui contiendra les tables sur votre ordinateur**. Après exécution de l'export sur le bureau à distance, copiez-collez les fichiers .sas et .dat dans le dossier en question.

Pour les dérivées 2 et 4 il faut également exporter via l'outil de requête de CSExport le fichier contenant l'ensemble des formats de l'eCRF à l'aide de la requête suivante :

```
SELECT AUTHORIZED_VALUES, CODE, VALUE_LG2
FROM BORD0016D.ASSEMBLY
INNER JOIN BORD0016D.LIBRARIES
ON AUTHORIZED_VALUES=CS_NAME
```

Le fichier obtenu doit être enregistré dans le dossier contenant les tables.

III. CONTRAINTES

Pour **toutes les dérivées** (contraintes du moteur) il faut :

- utiliser les fichiers SAS exportés par CSExport (le chemin et le nom de la bibliothèque doivent être modifiés avant l'export)
- avoir une variable VISIT contenant le nom des visites
- avoir une variable VISITNUM contenant le numéro de la visite
- avoir une variable USUBJID contenant l'identifiant du participant
- avoir une variable « --TESTCD » contenant les noms des futures variables.
- avoir une variable « --TEST » contenant les libellés des futures variables.

Pour les **dérivées 2 et 4** il faut, en plus des contraintes du moteur :

- avoir un fichier contenant tous les formats utilisés dans l'étude (obtenu à l'aide d'une requête dans CSExport)
- avoir une variable « --FMT » contenant le format de la valeur de « --ORRES » le cas échéant
- que la variable « --ORRES » soit préférentiellement au format numérique et ne contienne que des données numériques (pas de dates par exemple). Si cette contrainte n'est pas respectée, le programme fonctionne tout de même mais les formats sont appliqués en caractères et les données restent donc au format caractère.


IV. UTILISATION

Etape 1 : Copier le fichier « Croche-patte – v0.1 – 31052019 » dans le dossier de votre choix. Ce fichier est disponible sous :

R:\EQUIPES\USMR\DATA MANAGEMENT\Stagiaire\Pauline GAFFEZ\CDISC_SOS_TRIAL\Programmation\Croche-patte - v0.1 – 31052019

Etape 2 : Ouvrir et remplir les paramètres du fichier « Croche-patte – v0.1 - 31052019 ».

Noms des macros	Dérivées	Description
strChemIn	Toutes	Chemin du dossier dans lequel se trouvent les tables d'origine, les formats et où seront exportées les tables transposées et le rapport PDF.
Formats	2 et 4	Nom du fichier Excel contenant les formats (ex : RES_ORACLE_EXCEL_20190307_135941.xls).
Tab1-Tabn	Toutes	Noms des tables que l'on souhaite transposer. Le nom de ces tables est par la suite désigné par XX (ex : tab1=VS).

	Entité d'application : USMR Emetteur : USMR	IN-USM-
	INSTRUCTION	Ind : 0.1 Page : 3/4
Notice d'utilisation du programme Croche-patte v0.1		

Var_com	Toutes	Variables communes à toutes les tables que l'on souhaite garder une fois les tables transposées, séparées par des espaces (ex : « STUDYID USUBJID ARM ARMCD »). Si le module A est utilisé, les variables contenant le nom et numéro de la visite doivent y figurer.
XX_var1-XX_vark	Toutes	Noms des variables CDISC que l'on souhaite transposer (ex : VS_var1=VSORRESU). Ces paramètres sont à renseigner pour chaque table à traiter.
XX_s1-XX_sk	Toutes	Suffixes ajoutés aux noms des variables (noms définis par la variable « XXTESTCD ») lors de la transposition, commençant par un « _ ». Ces paramètres doivent obligatoirement être renseignés pour chaque variable CDISC définie précédemment (ex : VS_s1=_U).
XX_lab1-XX_labk	3 et 4	Précisions ajoutées aux labels des variables (labels définis par la variable « XXTEST »). Ces paramètres doivent obligatoirement être renseignés pour chaque variable CDISC définie précédemment (ex : VS_lab1=unité).
Module_A (ou B)	Toutes	Inclusion ou non des modules optionnels (1=oui 0=non). ⚠ Le module B permet de supprimer toutes les colonnes sans aucune donnée (colonnes construites par le programme). Il peut donc amener à supprimer des colonnes pour lesquelles aucune donnée n'est disponible.
Derivee	Toutes	Numéro de la dérivée que l'on souhaite lancer (1=export brut 2=formats 3=libellés 4=formats et libellés)

ETAPE 3 : Lancer le programme en cliquant sur Run.

Les tables obtenues sont disponibles sous SAS dans la librairie « Transpo » et sont enregistrées dans le dossier contenant les tables d'origine. Elles portent les noms T_XX, avec « XX » le nom d'origine. Le rapport PDF est également enregistré dans ce dossier.

Vous pouvez réitérer les étapes 2 et 3 pour lancer différentes dérivées et différents modules.

⚠ Cela écrasera les tables précédemment obtenues si elles n'ont pas été renommées. De même, le rapport PDF édité au premier lancement peut être écrasé selon la dérivée lancée et la date de la seconde exécution. Si vous lancez deux fois la même dérivée le même jour, seule la dernière version sera gardée (pour éviter cela il suffit de renommer le fichier PDF avant de lancer à nouveau le programme).

ANNEXE 8 : Protocole de qualification du programme Croche-patte

	Entité d'application : USMR Emetteur : USMR	EN-USM-803
	DOCUMENT D'ENREGISTREMENT	Ind : 0.1 Page : 1/3
	PROTOCOLE DE QUALIFICATION	

PROTOCOLE DE QUALIFICATION

Date : 23/05/2019

Programme : Croche-patte

Type de qualification :

Installation Opérationnelle Performance

Participants au groupe d'élaboration : Pauline GAFFEZ (data manager), Sophie CANETE (responsable management de la qualité), Guillaume DUPOUY (informaticien).

TABLEAU D'APPROBATION			
	REDACTION	VALIDATION	APPROBATION
Nom :			
Fonction :			
Date :			
Signature :			

	Entité d'application : USMR Emetteur : USMR	EN-USM-803
	DOCUMENT D'ENREGISTREMENT	Ind : 0.1 Page : 3/3
	PROTOCOLE DE QUALIFICATION	

- Exécution du script « SC02 - Suppression d'un espace.vbs »

6.3 Mauvais paramétrage des tables (tab1=TS)

- Exécution du fichier « Croche-patte_scenario3_v0.1_31052019.sas »
- Edition du rapport PDF avec notification des erreurs

6.4 Mauvais paramétrage du programme (derivee=5)


- Exécution du fichier « Croche-patte_scenario4_v0.1_31052019.sas »
- Message d'erreur dans la log et arrêt du programme

6.5 Jeu de données incorrect (pas de variable VISIT)

- Exécution du fichier « Croche-patte_scenario5_v0.1_31052019.sas »
- Edition du rapport PDF avec notification des erreurs

6.6 Variable --ORRES au format caractère

- Exécution du fichier « Croche-patte_scenario6_v0.1_31052019.sas »
- Edition du rapport PDF avec notification des erreurs

	Entité d'application : USMR Emetteur : USMR	EN-USM-803
	DOCUMENT D'ENREGISTREMENT	Ind : 0.1 Page : 2/3
	PROTOCOLE DE QUALIFICATION	

1 Descriptif du programme

Croche-patte est un programme développé en interne à l'USMR qui permet de transposer des tables initialement au format CDSIC pour obtenir des tables dans un format horizontal. Il est développé pour tourner sur des tables issues de CSEexport.

Le programme se décline en plusieurs dérivées qui suivent la transposition des tables :

- Export brut sans autre modification
- Application des formats récupérés dans CSEexport puis export
- Application de nouveaux libellés puis export
- Application des formats récupérés dans CSEexport et des nouveaux libellés puis export

La mise en œuvre du protocole de qualification donne lieu à un rapport de qualification (EN-USM-805).

2 Environnement du programme

SAS 9.4 (64 bits) sous Windows 10.

3 Tables sélectionnées

Les tables de référence utilisées sont des tables SAS au format CDISC construites pour l'étude SOStrial : VS, QS et DS avec les données des patients-test saisis. Elles sont issues du module CSEexport. Des tables légèrement modifiées seront utilisées dans les scénarios 5 et 6.

4 Prérequis

Le programme Croche-patte doit être déployé conformément à l'instruction en vigueur (IN-USM-REF).

Les prérequis sont de disposer du dossier contenant tous les fichiers du programme (avec les différents scénarios) ainsi que des tables SAS de référence.

5 Empreintes MD5 des modules du programme

Pour l'authentification du programme, l'empreinte MD5 des fichiers SAS est la suivante : 6DAFECF3F0E6E125D09C70CA37BD3913

6 Scénarios de qualification

6.1 Transposition et application des formats et libellés correctes

- Exécution du fichier « Croche-patte_scenario1_v0.1_31052019.sas »
- Edition du rapport PDF sans erreur notifiée

6.2 Authentification du programme grâce aux empreintes MD5

- Exécution du script « SC02 - Ajout d'un espace.vbs »
- Exécution du fichier « Croche-patte_scenario2_v0.1_31052019.sas »
- Message d'erreur dans la log et arrêt du programme

ANNEXE 9 : Rapport de qualification du programme Croche-patte

 CHU Bordeaux	Entité d'application : USMR Emetteur : USMR	EN-USM-805
	DOCUMENT D'ENREGISTREMENT	Ind : 02
	RAPPORT DE QUALIFICATION	Page : 1/8

RAPPORT DE QUALIFICATION

Date : 2019

Logiciel : Croche-patte v0.1 - 31052019

Type de qualification : Installation Opérationnelle Performance

Date du protocole de qualification : 23 mai 2019

Participants au groupe d'élaboration : Sophie CANETE (responsable management de la qualité), Guillaume DUPOUY (informaticien), Séverine MARTIREN (informaticienne).

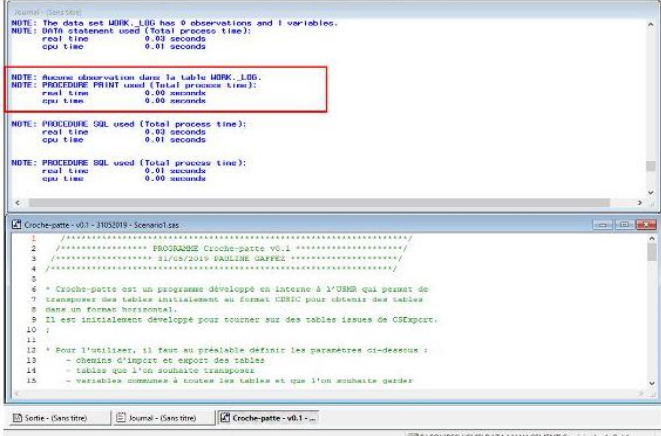
 CHU Bordeaux	Entité d'application : USMR Emetteur : USMR	EN-USM-805
	DOCUMENT D'ENREGISTREMENT	Ind : 02
	RAPPORT DE QUALIFICATION	Page : 2/8

1 Etapes de qualification

La mise en œuvre de la qualification s'est déroulée conformément aux étapes de qualification du protocole de qualification « EN-USM-803-Protocole de qualification Croche-patte - V0.1 – 31052019 ».

2 Résultats de qualification

2.1 Transposition et application des formats et libellés correctes



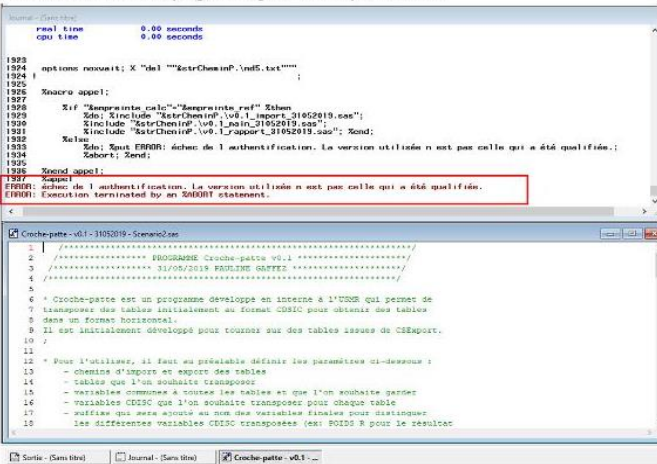
```

1 //***** PROLOGE Croche-patte v0.1 *****/
2 //***** 31/05/2019 PAULINE GAFFEZ *****/
3 //*****
4 //*****
5
6 * Croche-patte est un programme développé en interne à l'USMR qui permet de
7 transposer des tables initialement au format CDISC pour obtenir des tables
8 dans un format horizontal.
9 Il est initialement développé pour tourner sur des tables issues de CDExport.
10
11
12 * Pour l'utiliser, il faut au préalable définir les paramètres ci-dessous :
13 - chemins d'import et export des tables
14 - tables que l'on souhaite transposer
15 - variables CDISC que l'on souhaite transposer pour chaque table
16 - variables CDISC que l'on souhaite transposer pour chaque table
17 - suffixe qui sera ajouté au nom des variables finales pour distinguer
18 les différentes variables CDISC transposées (ex: P01S R pour le résultat)
  
```

La table work_log contient les erreurs éventuelles retrouvées lors de l'exécution du programme. Cette table est imprimée dans le rapport édité lorsqu'elle n'est pas vide.

 CHU Bordeaux	Entité d'application : USMR Emetteur : USMR	EN-USM-805
	DOCUMENT D'ENREGISTREMENT	Ind : 02
	RAPPORT DE QUALIFICATION	Page : 3/8

2.2 Authentification du programme grâce aux empreintes MD5

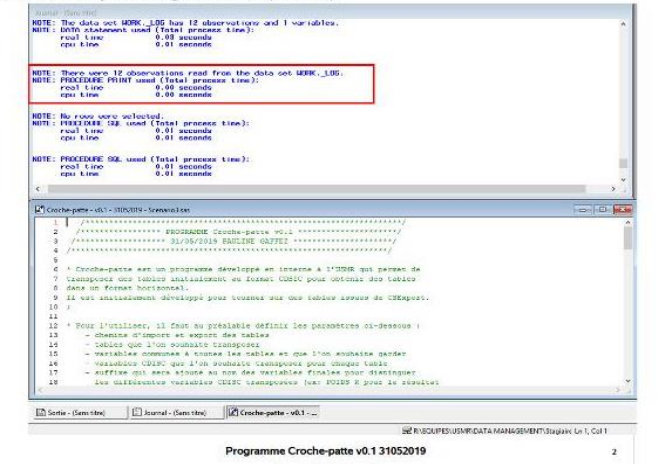


```

1 //***** PROLOGE Croche-patte v0.1 *****/
2 //***** 31/05/2019 PAULINE GAFFEZ *****/
3 //*****
4 //*****
5
6 * Croche-patte est un programme développé en interne à l'USMR qui permet de
7 transposer des tables initialement au format CDISC pour obtenir des tables
8 dans un format horizontal.
9 Il est initialement développé pour tourner sur des tables issues de CDExport.
10
11
12 * Pour l'utiliser, il faut au préalable définir les paramètres ci-dessous :
13 - chemins d'import et export des tables
14 - tables que l'on souhaite transposer
15 - variables CDISC que l'on souhaite transposer pour chaque table
16 - variables CDISC que l'on souhaite transposer pour chaque table
17 - suffixe qui sera ajouté au nom des variables finales pour distinguer
18 les différentes variables CDISC transposées (ex: P01S R pour le résultat)
  
```

 CHU Bordeaux	Entité d'application : USMR Emetteur : USMR	EN-USM-805
	DOCUMENT D'ENREGISTREMENT	Ind : 02
	RAPPORT DE QUALIFICATION	Page : 4/8

2.3 Mauvais paramétrage des tables (tab1=TS)



```

1 //***** PROLOGE Croche-patte v0.1 *****/
2 //***** 31/05/2019 PAULINE GAFFEZ *****/
3 //*****
4 //*****
5
6 * Croche-patte est un programme développé en interne à l'USMR qui permet de
7 transposer des tables initialement au format CDISC pour obtenir des tables
8 dans un format horizontal.
9 Il est initialement développé pour tourner sur des tables issues de CDExport.
10
11
12 * Pour l'utiliser, il faut au préalable définir les paramètres ci-dessous :
13 - chemins d'import et export des tables
14 - tables que l'on souhaite transposer
15 - variables CDISC que l'on souhaite transposer pour chaque table
16 - variables CDISC que l'on souhaite transposer pour chaque table
17 - suffixe qui sera ajouté au nom des variables finales pour distinguer
18 les différentes variables CDISC transposées (ex: P01S R pour le résultat)
  
```

Programme Croche-patte v0.1 31052019

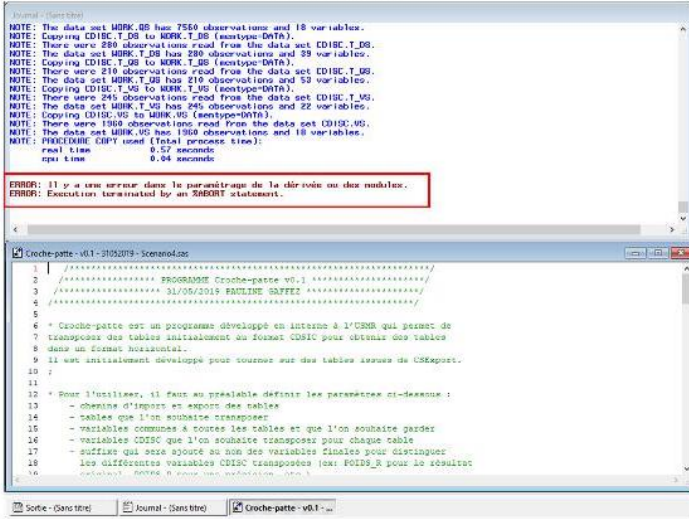
Rapport dérivé n°4

Application des formats, des libellés et export des tables transposées

Obs.	msg
1	ERROR: Le fichier WORK_TS.DATA n'existe pas.
2	ERROR: Le fichier WORK_LISTE_VISITE_TS.DATA n'existe pas.
3	ERROR: A character operand was found in the %EVAL function or %IF condition where a numeric operand
4	ERROR: The %TO value of the NDO IDX_VIS loop is invalid.
5	ERROR: Le fichier WORK_TS.DATA n'existe pas.
6	ERROR: A character operand was found in the %EVAL function or %IF condition where a numeric operand
7	ERROR: The %TO value of the NDO IDX_VIS loop is invalid.
8	ERROR: Le fichier WORK_TS_FMT.DATA n'existe pas.
9	ERROR: Le fichier WORK_TS_DATA n'existe pas.
10	ERROR: A character operand was found in the %EVAL function or %IF condition where a numeric operand
11	ERROR: The %TO value of the NDO IDX_FMT loop is invalid.
12	ERROR: Le fichier TRANSPOT_TS_DATA n'existe pas.

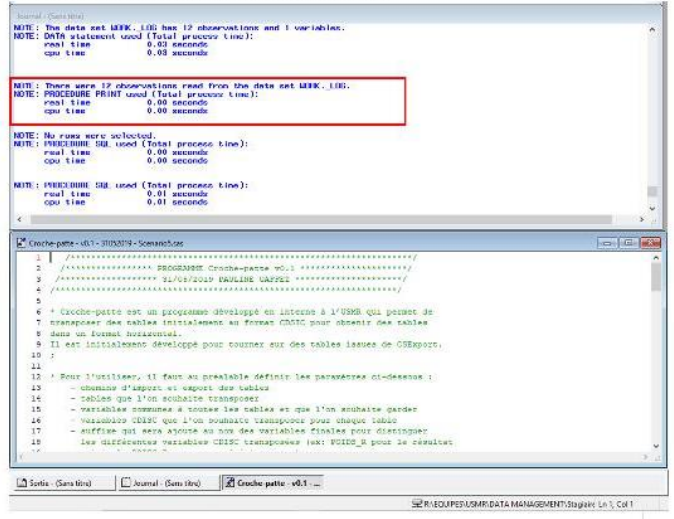
	Entité d'application : USMR	EN-USM-805
	Emetteur : USMR	
	DOCUMENT D'ENREGISTREMENT	Ind : 02 Page : 5/8
RAPPORT DE QUALIFICATION		

2.4 Mauvais paramétrage du programme (derivée=5)



	Entité d'application : USMR	EN-USM-805
	Emetteur : USMR	
	DOCUMENT D'ENREGISTREMENT	Ind : 02 Page : 6/8
RAPPORT DE QUALIFICATION		

2.5 Jeu de données incorrect (pas de variable VISIT)



Programme Croche-patte v0.1 31052019

2

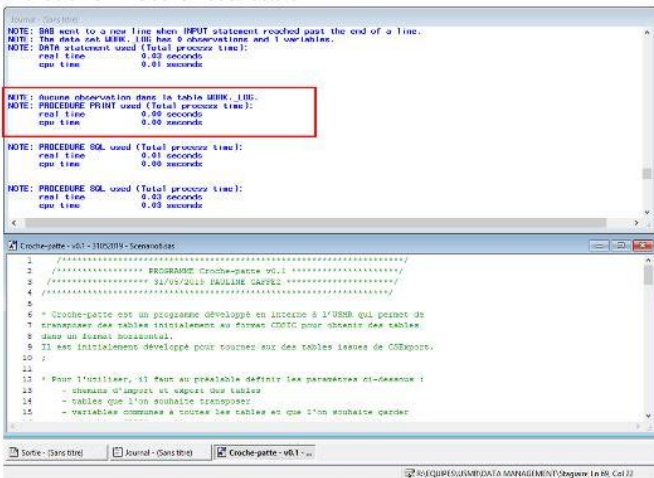
Rapport dérivée n°4

Application des formats, des libellés et export des tables transposées

Obs.	mLOG
1	ERROR: The following columns were not found in the contributing tables: VISIT.
2	ERROR: Le fichier WORK_LISTE_VISITE_VS_EX_DATA n'existe pas.
3	ERROR: A character operand was found in the %EVAL function or %IF condition where a numeric operand
4	ERROR: The NTO value of the %DO IDX_VS loop is invalid.
5	ERROR: The following columns were not found in the contributing tables: VISIT.
6	ERROR: A character operand was found in the %EVAL function or %IF condition where a numeric operand
7	ERROR: The NTO value of the %DO IDX_VS loop is invalid.
8	ERROR: Le fichier WORK_VS_EX_FMT DATA n'existe pas.
9	ERROR: Le fichier WORK_T_VS_EX_DATA n'existe pas.
10	ERROR: A character operand was found in the %EVAL function or %IF condition where a numeric operand
11	ERROR: The NTO value of the %DO IDX_FMT loop is invalid.
12	ERROR: Le fichier TRANSPRO_T_VS_EX_DATA n'existe pas.

	Entité d'application : USMR	EN-USM-805
	Emetteur : USMR	
	DOCUMENT D'ENREGISTREMENT	Ind : 02 Page : 7/8
RAPPORT DE QUALIFICATION		

2.6 Variable --ORRES au format caractère



3 Statut de qualification

Conforme

ANNEXE 10 : Script de l'intervention en forum méthodologique du 12/06/2019

Présentation de stage : contexte, objectif et avancement

« *Standardisation via CDISC : Intérêts et limites pour le data management et la réalisation d'analyses statistiques à l'USMR* »

DIAPO 1

Brève présentation

DIAPO 2

La gestion de données biomédicales a pour objectif de fournir des données de qualité pour les analyses. Ces dernières années elle a évolué en réponse au passage de cahiers d'observation papier à des cahiers électroniques. Ces nouvelles pratiques nécessitent de standardiser les procédures et les terminologies utilisées pour améliorer la qualité des données et gagner en efficacité.

La standardisation est « le processus de parvenir à un accord sur les définitions communes des données, les formats, la représentation et les structures de toutes les couches de données et éléments » (OCDE).

DIAPO 3

Dans la recherche, les standards fournissent une connaissance commune qui définit comment échanger les données mais aussi comment les collecter, les comprendre et les utiliser. Cela permet d'optimiser la gestion de données par le recours à une structure unique et donc de faciliter les échanges d'informations et l'intégration de différentes sources de données dans des registres ou des entrepôts par exemple. Cette structure permet également de faciliter les méta-analyses en agrégeant des données provenant de diverses études précédemment menées. Enfin, les standards peuvent permettre la soumission à différentes autorités telles que la FDA qui impose le format CDISC pour l'obtention d'une autorisation de mise sur le marché.

DIAPO 4

Le CDISC (*Clinical Data Interchange Standards Consortium*) est un consortium international à but non lucratif qui cherche à promouvoir la standardisation des formats de recueil, d'échange, et de soumission de données dans la recherche clinique. Le but est de fluidifier la circulation des données depuis leur récolte jusqu'à leur analyse afin de réduire les temps de traitement et donc les coûts. Pour ce faire, CDISC a créé et publié une série de standards qui sont régulièrement mis à jour et qui interviennent à différentes étapes de la recherche (cf image). Le standard auquel je m'intéresse plus particulièrement est le SDTM qui permet la structuration des tables de données.

DIAPO 5

Le *Study Data Tabulation Model* (SDTM) permet la présentation standardisée des données des études cliniques. Sa structure repose sous la forme de tables appelées domaines et regroupant des observations sur un même thème (exemple « VS »).

Ce qu'il faut bien comprendre dans le SDTM c'est le passage d'une structure horizontale, c'est-à-dire qu'une ligne correspond à l'ensemble des observations pour un patient, à une structure verticale dans laquelle une ligne correspond à une seule observation (contenant plusieurs informations telles que la mesure, l'unité et la date par exemple) d'un patient. Cf. exemple.

DIAPO 6

CDISC a créé des guides définissant des normes pour permettre la mise en place de ses standards dans différentes aires thérapeutiques. Il existe à ce jour 31 guides publiés, notamment pour Alzheimer, le cancer du sein ou le diabète. Ces guides permettent de standardiser le recueil de données dans un champ thérapeutique précis.

Des exemples de recours au CDISC pour différents projets se trouvent dans la littérature. Par exemple à l'aide du SDTM, l'équipe du data center UNICANCER a pu définir les domaines à utiliser et ils ont défini des CRF-types en fonction des contraintes de leurs études. Pour chaque CRF-type, des tables décrivant les métadonnées ont été créées et des listes de codes issues des terminologies CDISC ont été dressées. Cela a permis de créer une structure reproductible qui permet une simplification de la mise en commun des données des essais pour faciliter les méta-analyses.

Un autre exemple a été publié en 2014. Matsumara et son équipe sont parvenus à connecter des eCRF à des systèmes d'enregistrements médicaux électroniques à l'aide du CDISC. Cela a permis de remplir automatiquement les eCRF et donc de ne pas avoir à transcrire manuellement les données lues dans un dossier médical ou sur un cahier d'observation papier.

DIAPO 7

Un autre standard se développe depuis quelques années, en particulier dans le domaine du soin : HL7 FHIR. C'est un standard facile à appliquer et souple fonctionnant sur le principe de ressources. Ces ressources couvrent 80% des données récoltées, les 20% restant trouvent leur place dans les extensions.

La possibilité de fusionner CDISC et HL7 FHIR a été étudiée dans la littérature. Certains ont cherché à créer des *mappings* entre CDISC et HL7 FHIR, d'autres se sont penchés sur le standard CDISC ODM v2.

Tous les articles s'accordent sur le fait qu'il est nécessaire que ces deux standards convergent dans un futur proche, car les données du soin pourraient servir à des fins de recherche à l'avenir.

DIAPO 8

Mon stage consiste à explorer la faisabilité et l'intérêt de la mise en place du CDISC à l'USMR pour la gestion de données. Pour ce faire, je reprends le projet R&D entamé par Olivier Quintin qui était de mettre en place le CDISC sur l'étude SOStrial qui est gérée sous Ennov. L'étude SOStrial PHRC national dont l'investigateur principal est le Professeur Jean-François Dartigues. L'objectif principal est la comparaison de deux stratégies thérapeutiques chez des patients présentant une maladie d'Alzheimer légère à modérée. Suite au déremboursement du médicament par l'Assurance Maladie l'étude a été stoppée, avec une possibilité de reprise si un financement est trouvé.

Les objectifs de mon stage sont donc :

- D'explorer les avantages et limites du CDISC
- D'étudier la faisabilité de sa mise en place et l'intérêt au sein de l'USMR
- De créer un programme sous SAS pour transposer les tables CDISC et les rendre exploitables par les programmes d'analyse standards
- D'élaborer des documents qualité si besoin

DIAPO 9

Ce qui a déjà été fait :

- Création d'un SDTM pour l'étude SOStrial et discussions avec Olivier Quintin et le groupe DM
- Création des tables d'export sous CSEXPORT à partir de l'eCRF désigné par Olivier
- Définition d'un cahier des charges avec Rémi Sitta (ce cahier est en cours de validation)
- Gros travail de programmation sous SAS pour transposer les tables automatiquement et gérer les formats
- Optimisation du programme (modularité notamment)
- Rédaction d'une notice d'utilisation

DIAPO 10

Les premières conclusions de ce travail sont qu'il est difficile de mettre en place du pur CDISC au sein de l'USMR à cause de la multitude de sujets traités et qu'il est plus faisable de standardiser en s'inspirant du CDISC, sur le principe du question/réponse. En revanche dans la conjoncture actuelle avec la refonte des circuits de données à l'USMR, le choix a été fait de standardiser les pratiques entre ENNOV et REDCap. Comme REDCap n'est pas nativement compatible avec le CDISC, il est peu probable que cela soit mis en place dans un futur proche.

Mais le programme développé pourrait tout de même être utile aux statisticiens, dans la mesure où il arrive qu'ils reçoivent des données au format CDISC. C'est pourquoi le travail a été poursuivi.

Maintenant il reste à finaliser le protocole de qualification du programme pour vérifier sa fiabilité afin de le faire valider. Et de finaliser également les différents documents qualité dans le SMQ.

En parallèle de tout cela il faudra que je poursuive la rédaction de mon mémoire pour pouvoir valider mon master.

RESUME

Le *Clinical Data Interchange Standards Consortium* (CDISC) est un consortium international à but non lucratif qui a publié une série de standard afin d'harmoniser les formats de recueil et d'échange de données tout au long d'une étude de recherche clinique. Le but de ce travail est de voir dans quelle mesure leur application est possible au sein d'une structure telle que l'Unité de Soutien Méthodologique à la Recherche clinique et épidémiologique (USMR) et si leur mise en place est nécessaire. La conclusion de ces réflexions est qu'il pourrait être envisagé de mettre en place les standards CDISC à l'USMR si les moyens étaient donnés à l'équipe, mais que leur application n'est pas nécessaire au vu des pratiques actuelles de l'unité. En revanche la standardisation des pratiques au sein de l'équipe l'est, et le système de management de la qualité permet de la mettre en œuvre. Afin d'assurer la continuité des activités de l'USMR, en particulier l'analyse des données, un travail de programmation a également été intégré à ce projet. Le programme développé permet de transposer des tables d'un format vertical vers un format horizontal.

Mots-clés : standardisation, recherche clinique, CDISC, gestion de données

ABSTRACT

Clinical Data Interchange Standards Consortium (CDISC) is a non-profit international consortium that has published a series of standards in order to harmonize the formats of data collection and exchange throughout a clinical research. The purpose of this work is to explore to what extent their application is possible within a structure such as the unit of methodological support to clinical and epidemiological research (USMR) and if their implementation is needed. The conclusion of these reflections is that it could be envisaged to implement the CDISC standards to USMR if resources were given to the team, but that their application is not necessary with the current practices of the unit. However standardization within the team is a necessity, and the quality management system enable to implement it. To ensure the continuity of USMR activities, especially data analysis, programming work has also been incorporated into this project. The developed program can convert tables in a vertical format to a horizontal format.

Keywords : standardization, clinical research, CDISC, data management