

Description des patients et de la prise en charge de la maladie chronique du greffon contre l'hôte contre l'hôte (cGVHD) en France basée sur le traitement du langage naturel (NLP) à partir des comptes-rendus médicaux électroniques de 8 centres en France

Colas S¹, Chantepie S^{2,6}, Daguindau E^{3,6}, Huynh A^{4,6}, Loschi M^{5,6}, Nguyen S⁶, Devillier R⁶, Raus N⁶, Boumendil A⁶, Requillard C⁷, Chuttoo L⁷, Poplu A⁷, Allali N¹, Kiprijanovski D¹, Le Ouay F⁷, Jeanbat V⁸, Kirion J⁸, Buchbinder N⁹, François S¹⁰, Villate A¹¹, Castilla-Llorente C¹², and Bouée S⁸

- 1 Sanofi, Paris, France
- 5 CHU de Nice, France
- 9 CHU de Rouen, France
- 10 CHU d'Angers, France
- 2 CHU de Caen, France
- 6 SFGM-TC, Paris, France
- 11 CHU de Tours, France
- 3 CHU de Besançon, France
- 7 Lifem, Paris, France
- 12 Institut Gustave-Roussy, Villejuif, France
- 4 Oncopôle de Toulouse, France
- 8 CEMKA, Bourg-La-Reine, France

INTRODUCTION

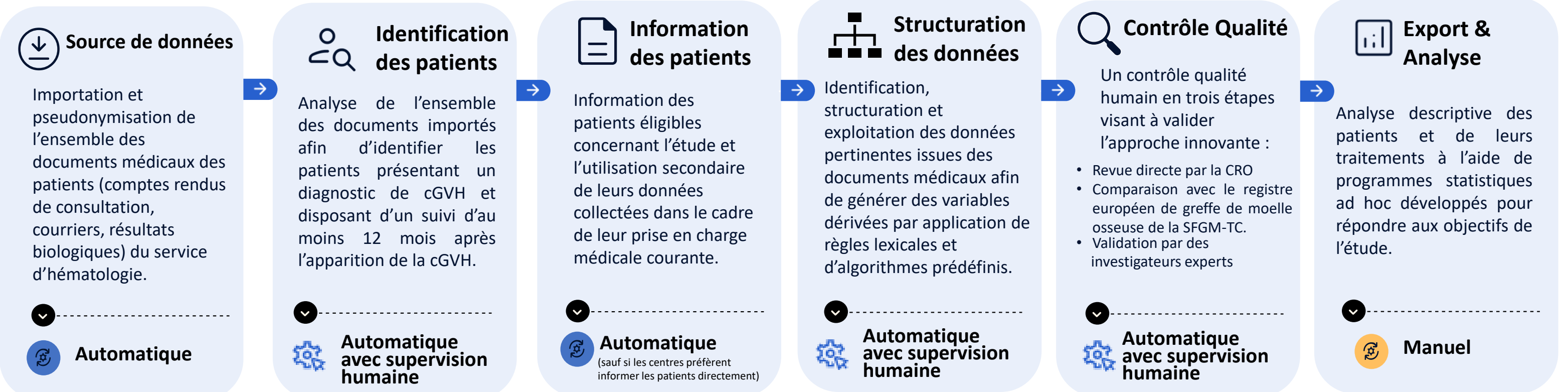
- ✓ La maladie chronique du greffon contre l'hôte (cGVHD) est l'une des complications les plus fréquentes et les plus graves après une allogreffe de cellules souches hématopoïétiques (allo-HSCT).
- ✓ En Europe, il n'existe pas de consensus thérapeutique pour les patients en échec de la deuxième ligne de traitement systémique (LOT) et la prise en charge thérapeutique au-delà de la première LOT est peu décrite.

OBJECTIFS

- ✓ Les objectifs étaient d'identifier les patients atteints de cGVHD, décrire leurs caractéristiques et les stratégies de traitements, via une méthode d'intelligence artificielle fondée sur le traitement automatique du langage naturel (IA-NLP) pour structurer les données issues des dossiers médicaux.

METHODES

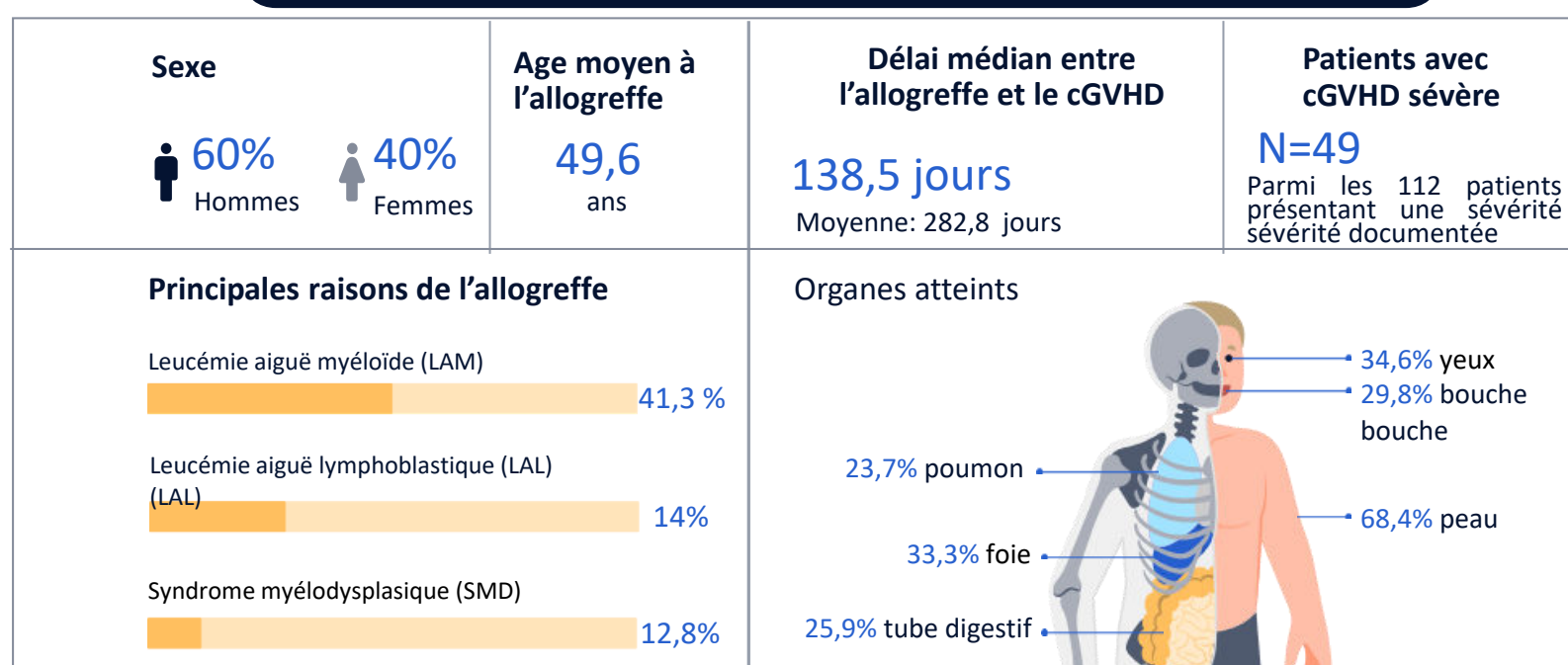
La spécificité de ce projet réside dans son approche innovante pour l'identification des patients et la constitution de la cohorte. Celle-ci repose sur une structuration automatisée des données fondée sur des méthodes d'intelligence artificielle de traitement du langage naturel (IA-NLP). L'AI-NLP a analysé les dossiers médicaux des patients ayant reçu une allo-HSCT dans 8 hôpitaux français. Puis, les données pseudonymisées des patients atteints de cGVHD suivis pendant au moins 1 an après apparition de la cGVHD ont été extraites et structurées. Ce qui a permis de décrire les caractéristiques des patients, de leur cGVHD, leurs traitements et les réponses aux traitements.



RESULTATS

Parmi les dossiers médicaux de 513 patients ayant bénéficié d'une allogreffe de cellules souches hématopoïétiques (allo-HSCT) issus des 8 établissements hospitaliers et examinés, 235 patients répondaient aux critères d'éligibilité (traités pour une cGVH), dont les données ont été extraites et structurées via les outils d'IA-NLP. Parmi ces 235 patients, 41 ont eu au moins 3 lignes de traitement (LDT).

Figure 1 : Description des patients et des traitements



Évaluation de la performance de l'IA générative

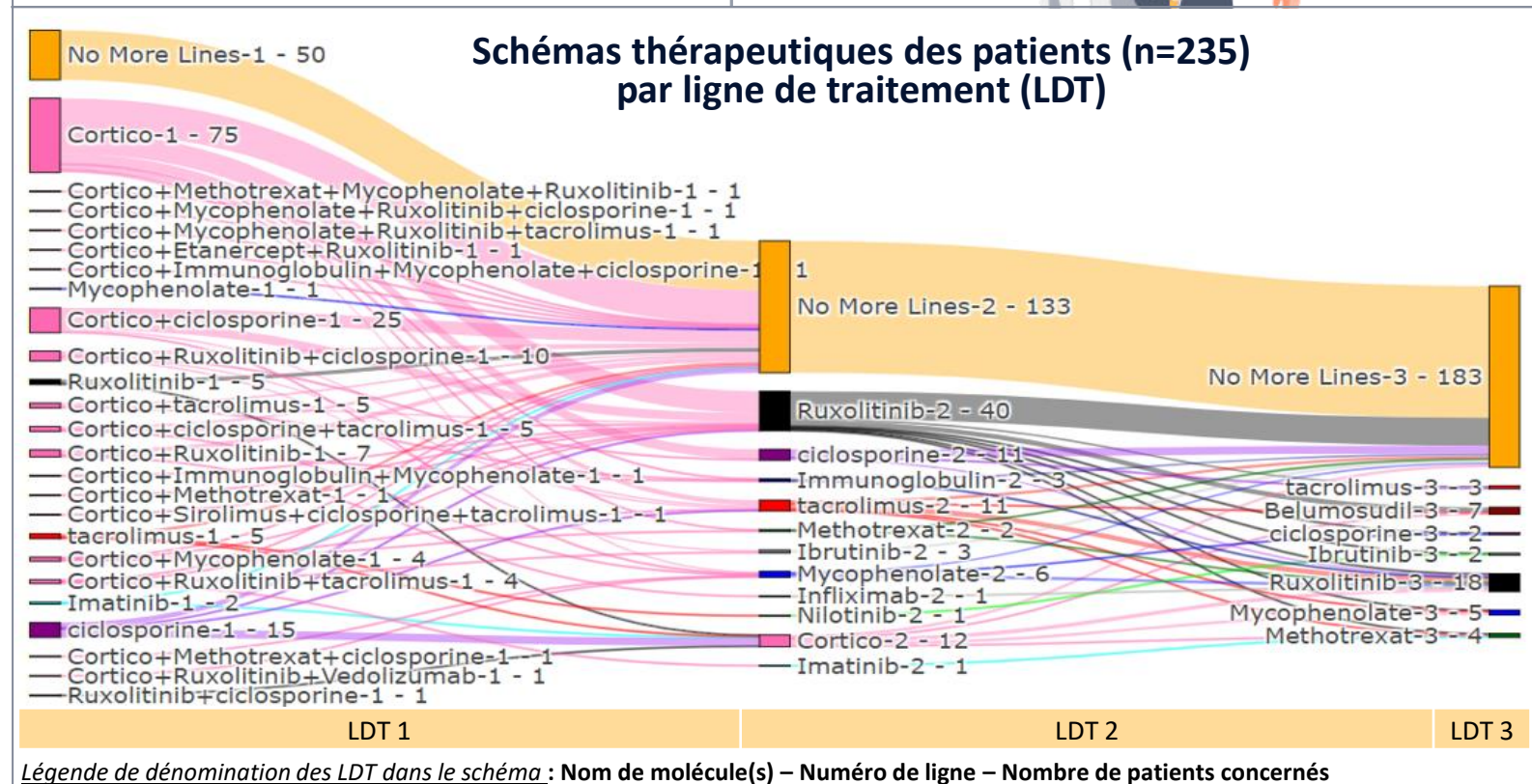
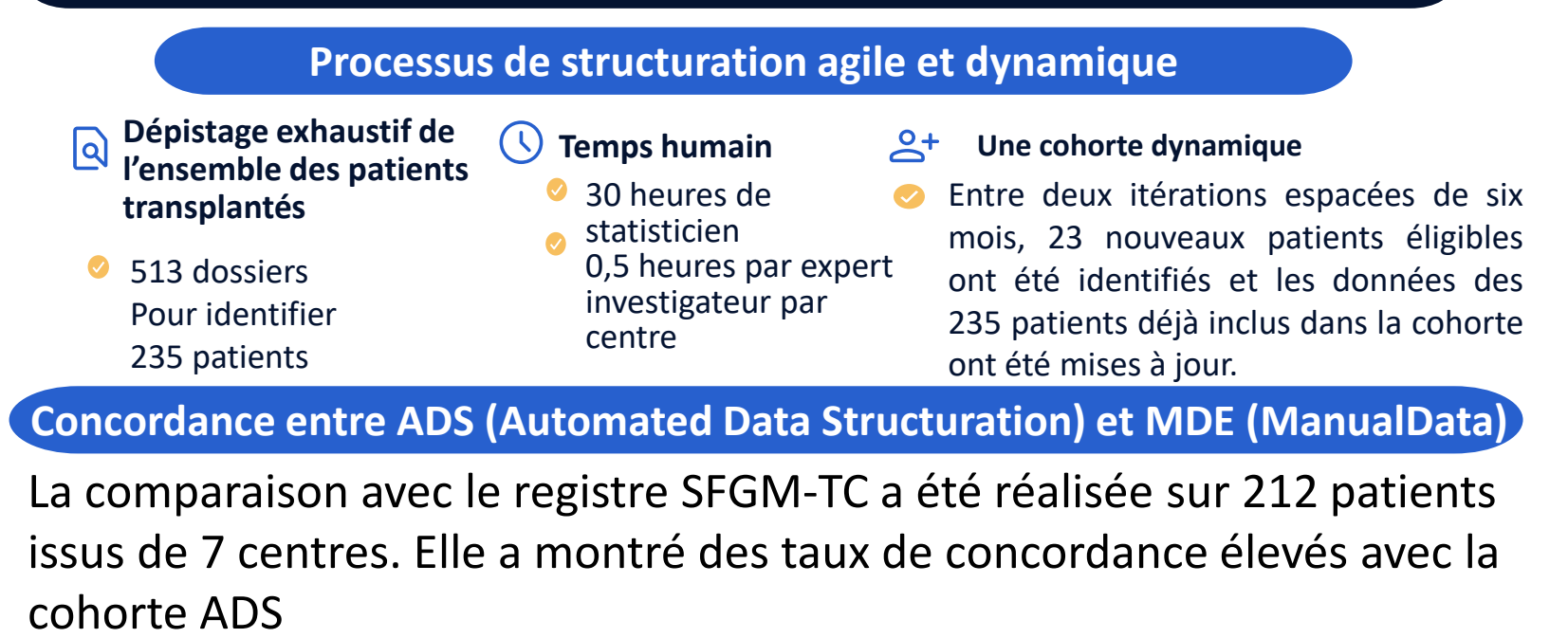
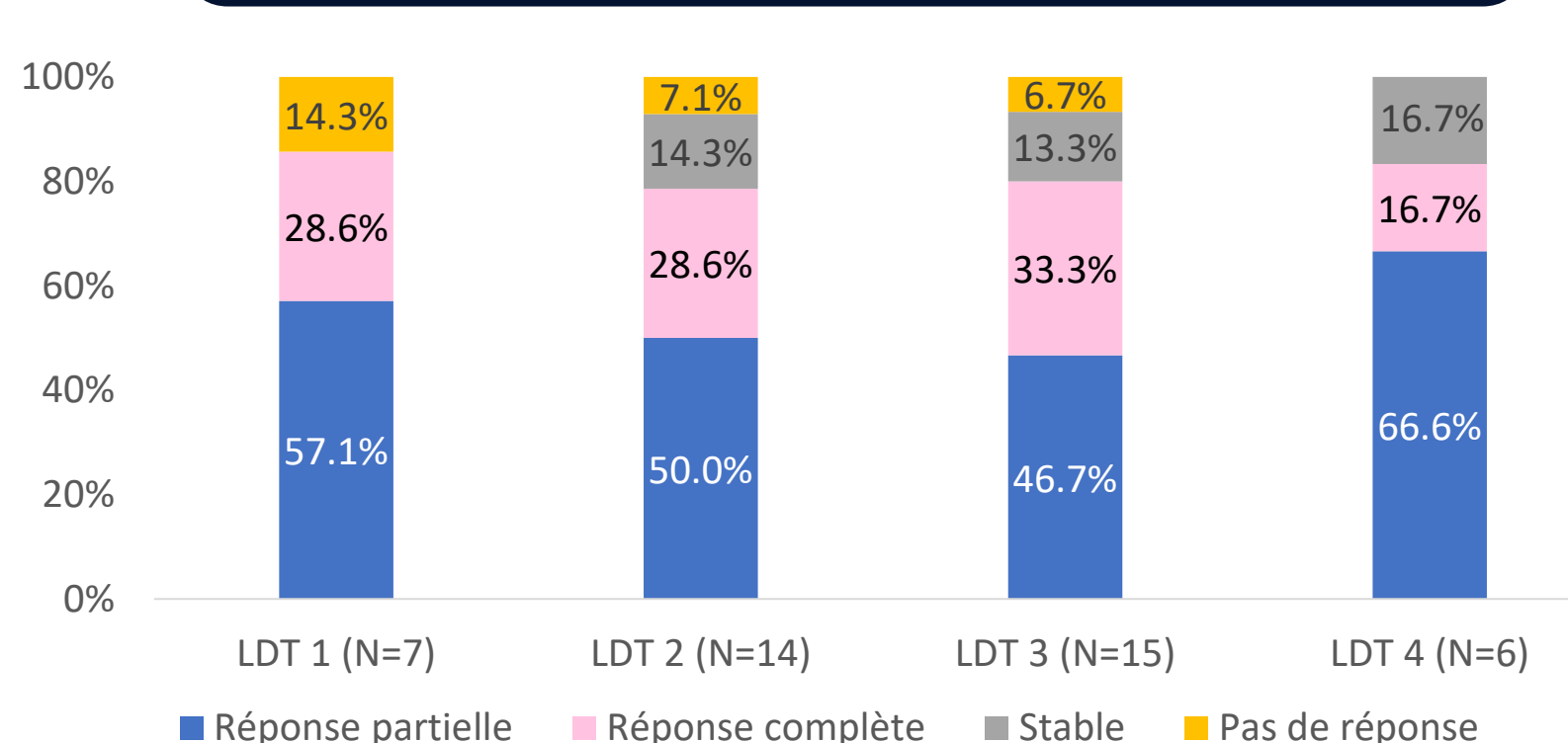


Figure 2 : Réponses au traitement par ligne de traitement



Vérifications des discordances

Parmi les 212 patients inclus dans la cohorte ADS, 204 ont été identifiés dans le registre SFGM-TC. Parmi les 29 variables (5 916 points de données) communes aux deux sources de données, 7 (les noms des essais cliniques, l'hôpital de réalisation de l'allo-HSCT, le poids, les comorbidités, les scores de Karnofsky et les organes atteints) présentaient un certain degré de discordance (n = 2 520 points discordants) de différents types.

- **Divergences ADS-MDE** liées à des champs à choix multiples dans les deux sources, entraînant un appariement partiel des valeurs (n=1 093 ; 43%)
- **Incohérences** dues à un accès limité aux données ADS avec un **historique de 2 ans** et des **données manquantes dans les dossiers médicaux du fait des pratiques médicales** (n=430 ; 17%)
- Incohérences dues à des informations obsolètes dans le MDE, à des données manquantes dans le registre SFGM-TC ou à d'autres erreurs.

En troisième ligne de traitement (Figure 2.)

- La réponse globale était renseignée pour 15 patients : 5 (33,3%) avaient une réponse complète, 7 (46,7%) une réponse partielle.
- Le délai moyen entre le début de la LOT et la 1ère réponse partielle ou complète était 91,7 jours (11 patients).
- La survie moyenne sans progression (délai entre le début de la LOT et le début d'un nouvel événement (nouveau traitement contre la cGVHD, rechute de la maladie sous-jacente ou décès) était 319,9 jours (14 patients).

CONCLUSION

- ✓ Ce processus innovant d'IA-NLP permet d'identifier et d'extraire facilement les données d'intérêt et de générer, avec une intervention humaine minimale, un jeu de données structuré robuste et précis incluant l'ensemble des patients atteints de cGVH dans les centres participants, à partir d'informations dispersées dans les dossiers médicaux et complexes à analyser manuellement.

- ✓ 8 établissements hospitaliers participent au projet et plus de 200 patients ont été inclus, avec des données désormais structurées. Ces premières étapes ont permis de valider l'outil d'IA-NLP et d'obtenir une description plus large des patients et de leurs prises en charge thérapeutiques.
- ✓ Les premières analyses descriptives, réalisées dans huit établissements hospitaliers et portant sur 235 patients, apportent un éclairage sur la prise en charge en conditions réelles de la cGVH en France.

DPI SCo et DK sont des employés de Sanofi. Sch, AH, ML, sont membres de la SFGM-TC qui a reçu un financement de Sanofi pour réaliser les analyses.

ED est membre de la SFGM-TC et a été dédommagé pour des trajets par Novartis et Neovii. FLO et AP sont des employés de Lifem qui effectuent l'extraction et la structuration des données pour l'étude mentionnée ci-dessus, financée par Sanofi. JK et SB sont employés par Cemka, qui a reçu une subvention de Sanofi pour réaliser les analyses de l'étude.

Contact : sandrine.colas@sanofi.com