

Objectif

Décrire de manière automatique les lignes de traitements reçues par les patients atteints d'un Myélome Multiple (MM) en France à partir du SNDS.

Contexte

- Le MM est une hémopathie maligne qui touche environ 5 000 nouveaux patients par an.
- Au cours de cette maladie se succèdent des périodes de rémissions et de rechutes, prises en charge par des protocoles de soins constitués de combinaisons de chimiothérapies et/ou greffe de cellules souche hématopoïétiques. Ces protocoles sont constitués d'une ou plusieurs phases. Chaque phase étant une répétition de cycles.
- À chaque nouvelle ligne de traitement, le protocole de soin est adapté par les cliniciens au regard du profil du patient et/ou des effets secondaires.
- Les données du SNDS, couplées à l'algorithme ATLAS, permettent d'identifier ces lignes de traitement.

Challenges

Détection et automatisation complexes

Théorie	Pratique	Data
Protocoles complexes (plusieurs médicaments combinés, avec des timings différents)	Déviations par rapport aux protocoles théoriques (adaptation des traitements, effets indésirables, etc.)	Seuls les traitements remboursés sont visibles
Protocoles nombreux	Des milliers de patients	Traitements intra-GHS non nommés
		Essais cliniques

Méthode : ATLAS, un algorithme d'intelligence artificielle en 8 étapes

1 Recensement de tous les cycles constituant les protocoles théoriques, avec l'aide des cliniciens
41 cycles recensés pour 26 protocoles théoriques

Cycle : séquence finie dans le temps, comprenant l'administration d'une ou plusieurs molécules suivant un rythme prédéfini

2 Sélection de la cohorte de patients atteints d'un MM à partir du SNDS, recherche de l'historique médical de chaque patient (médicaments, injections et actes composant les cycles théoriques)
17 442 patients incidents (2014 - 2017)

Pour chaque patient :

3 IA Alignement des données médico-administratives du patient avec les cycles théoriques recensés

4 IA Calcul d'un score de similitude pour chacun de ces alignements
Utilisation d'une adaptation de l'algorithme d'alignement de séquences ADN de Smith-Waterman

5 Visualisation des scores de similitudes retrouvés au cours du temps

6 IA Lissage des scores obtenus pour définir des phases

Phase : succession de cycles identiques au cours du temps

Utilisation d'une estimation de densité à noyau gaussien sur chaque cycle, afin de lisser les scores dans le temps

7 Passages des phases aux lignes

Ligne : succession d'une ou plusieurs phases formant une entité thérapeutique complète (un protocole de soin)

Utilisation de la connaissance médicale des experts cliniciens

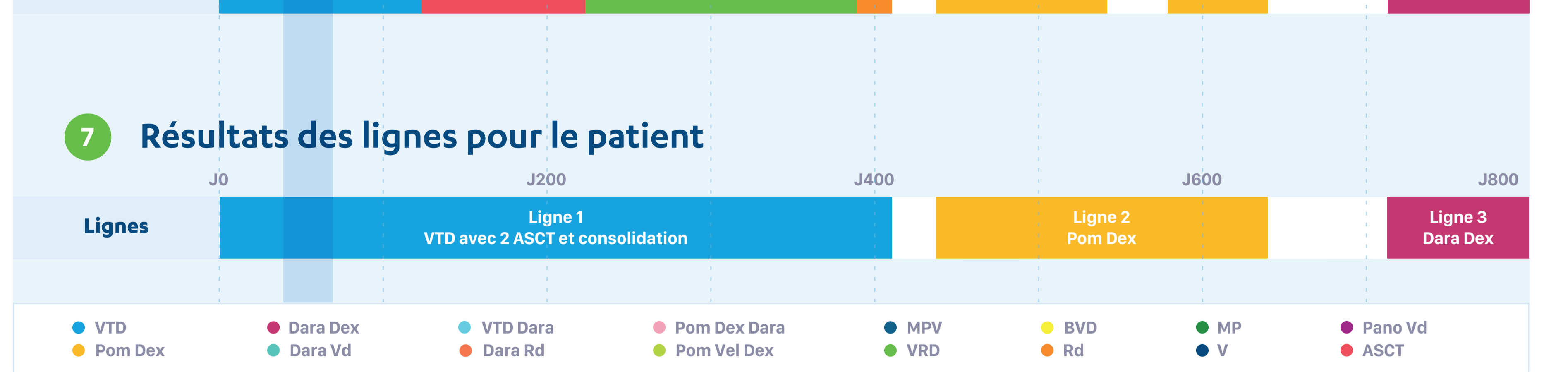
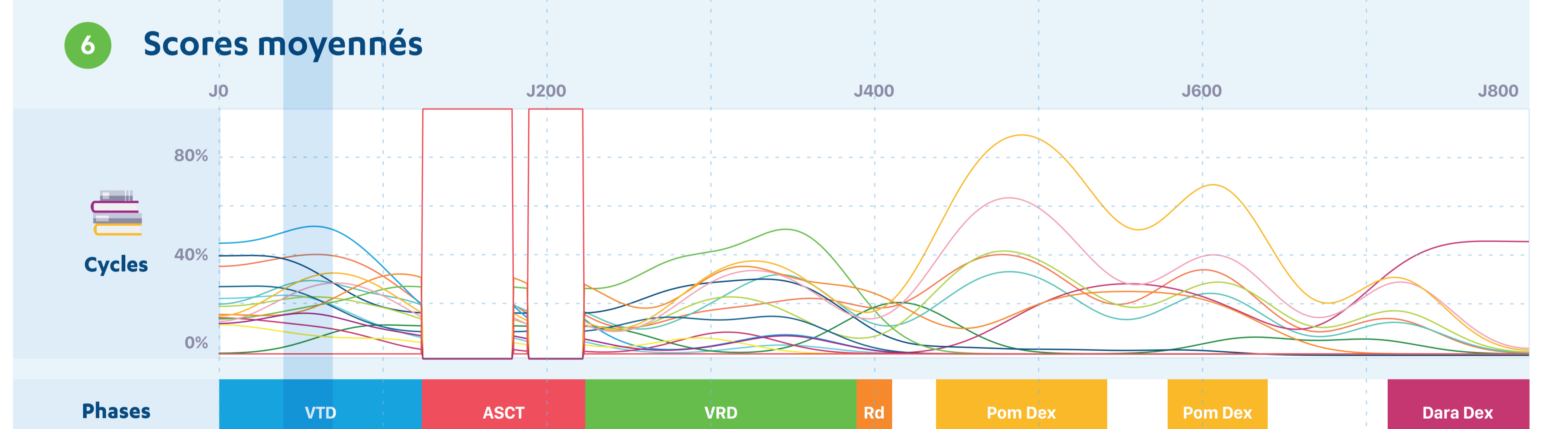
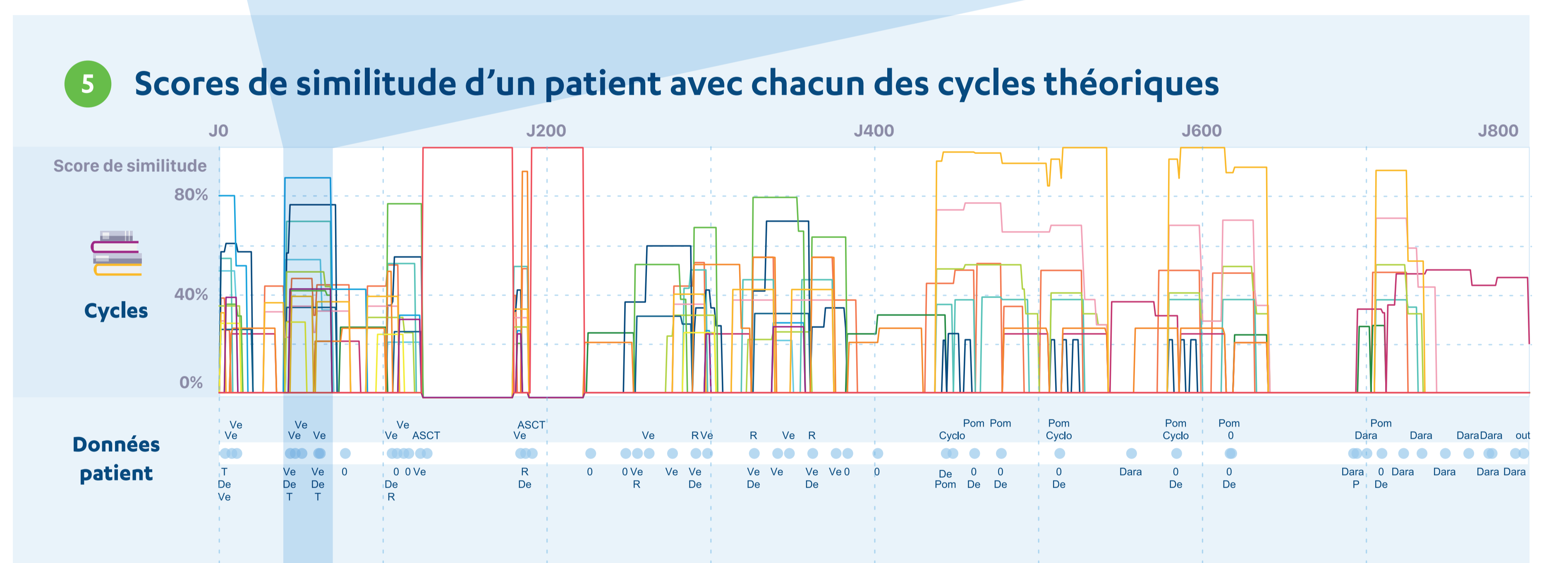
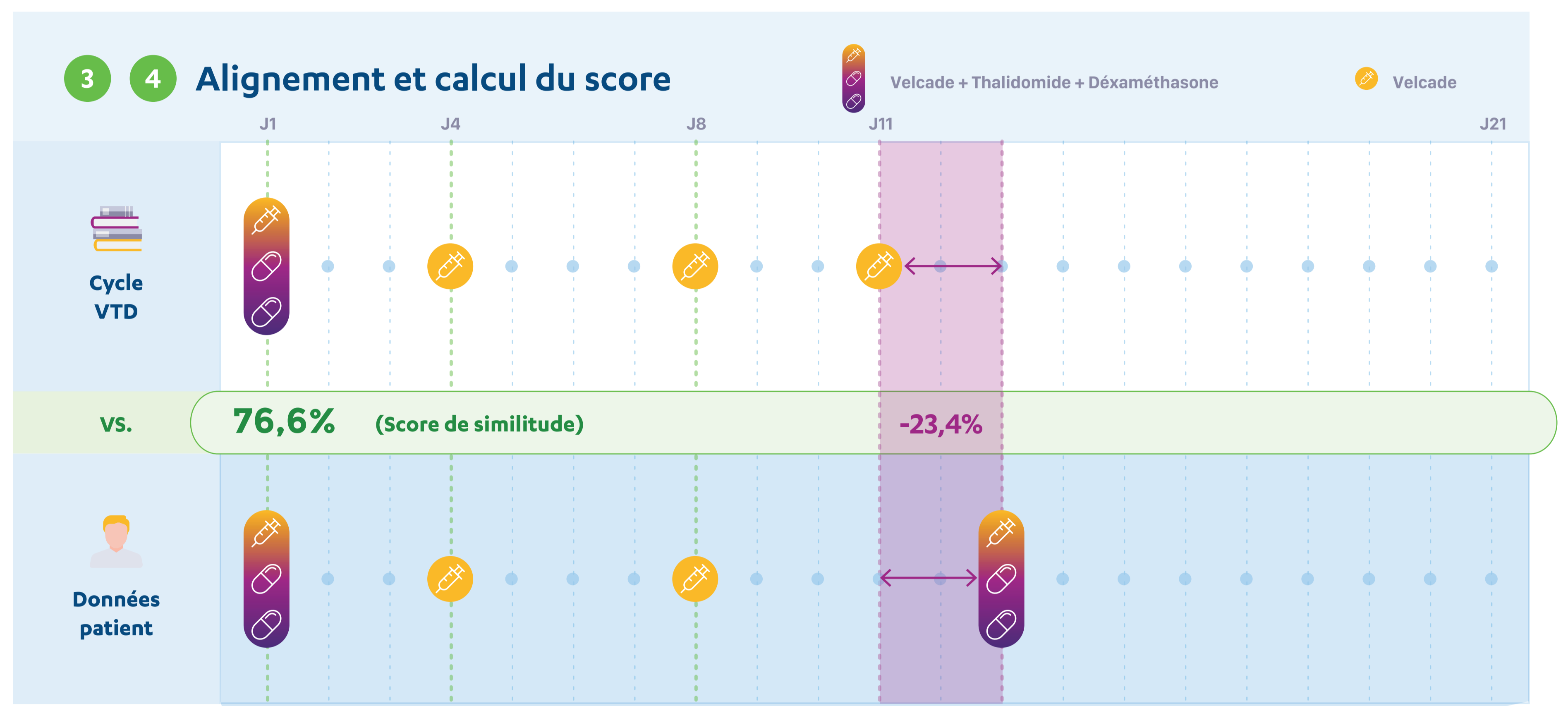
Exemple : des cycles VTD peuvent former l'induction pour préparer une greffe (ASCT), suivis de cycles VRD et Rd.

Sur l'ensemble de la cohorte :

8 Agrégation des résultats

Les étapes 3 à 7 se font indépendamment pour chaque patient. Les résultats sont finalement agrégés au niveau de la cohorte entière (histogrammes, distributions, Kaplan-Meier).

L'ensemble a été réitéré 2 fois, afin d'améliorer chaque étape pour qu'elle soit adaptée à la prise en charge réelle, afin que les résultats finaux soient le plus précis possible



Conclusion

- Méthode innovante**
Algorithme d'alignement de séquence d'ADN adapté à la recherche de séquences de soin
Technique de lissage gaussien
- Verrous scientifiques levés**
Flexibilité et résistance aux atypies
Automatisation
- Application médicale**
Vision complète des protocoles de traitement dans le domaine du myélome multiple (publications à venir)

L'épidémiologie de la prise en charge du MM peut être décrite, ouvrant la possibilité d'une meilleure description des séquences thérapeutiques sur un grand nombre de données patients en vie réelle
Cette méthode peut également s'appliquer à la description des lignes thérapeutiques d'autres pathologies

Résultats

1ère exécution de l'algorithme

77 % des patients ont leurs lignes totalement identifiées

- Ajout des 2 cycles identifiés à posteriori **1**
- Amélioration du paramétrage **3 4 6**
- Amélioration de la compréhension médicale **7**

2ème exécution de l'algorithme

92% des patients ont leurs lignes totalement identifiées
8% des patients ont des séquences trop distinctes des cycles théoriques pour être réidentifiées.