

Rapport Tache Technique

POSTE DE DATA SCIENTIST AP-HP / BNDMR

Données sur les patients atteints d'Epidermolyse bulleuse (EB)

Parmi les types et sous-types d'EB les patients sont équitablement distribués (Fig.1). De plus, il y a un nombre similaire de patients dans la base pour chaque grand type d'EB, respectivement 82 pour l'EB simple (EBS) et 76 pour l'EB jonctionnelle (EBJ), lorsque les diagnostics sont regroupés selon le schéma hiérarchique transmis (Fig.3).

Connaissant la date de naissance des patients, nous les avons regroupés en classes d'âge. Nous pouvons constater une distribution normale de l'âge des patients, et une distribution équitable entre l'EB simple et l'EB jonctionnelle pour chaque classe d'âge (Fig.2).

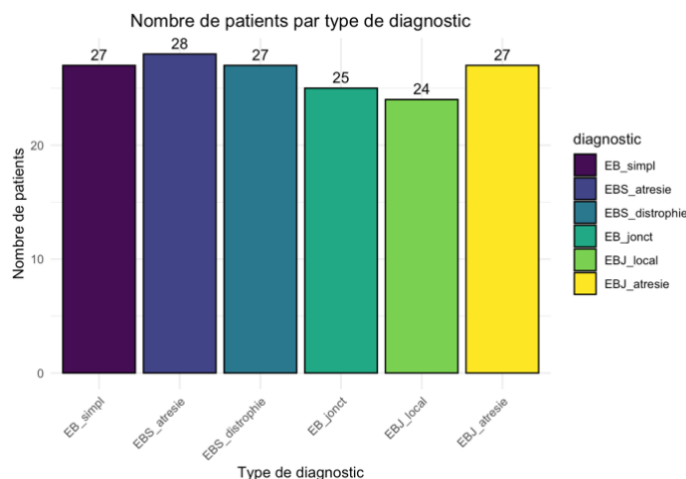
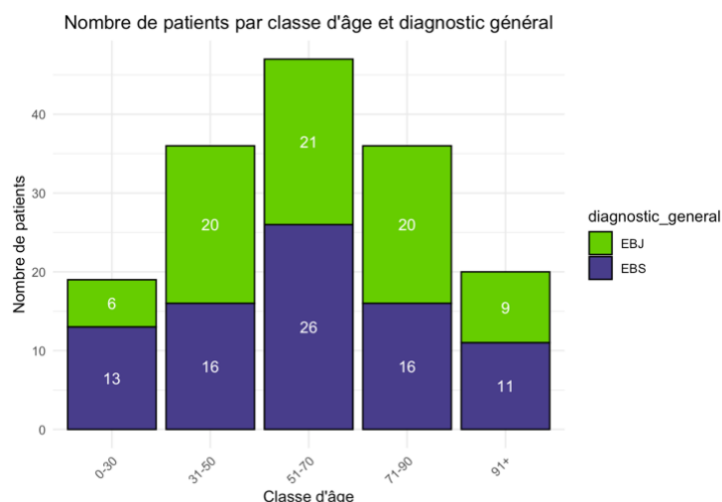


Figure 1: Répartition détaillée des type d'EB dans les patients.



chaque type d'EB selon leur classe d'âge.

Figure 2: Répartition des patients dans

Nous avons aussi voulu vérifier si le niveau de diagnostic et notamment le nombre de cas inconnus ou à l'inverse confirmés était similaire entre les 2 types de maladie et nous avons constaté que la distribution était quasiment identique (Fig.3). De même, nous avons encodé le temps de suivi à partir de la date d'inclusion et séparé en classes, et nous avons pu constater que cela n'avait pas d'effet sur le niveau de diagnostic des patients (Fig.4).

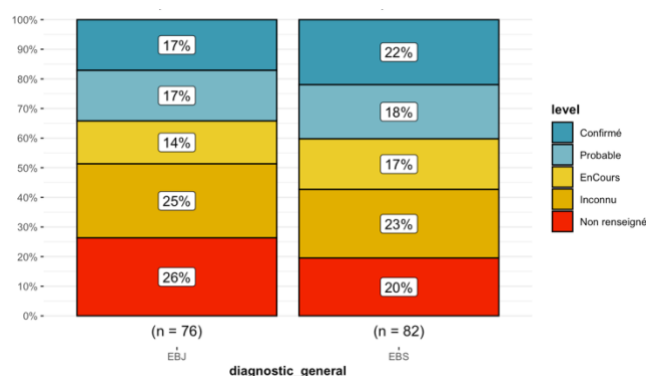


Figure 3: Distribution des niveaux de diagnostic pour les deux types d'EB.

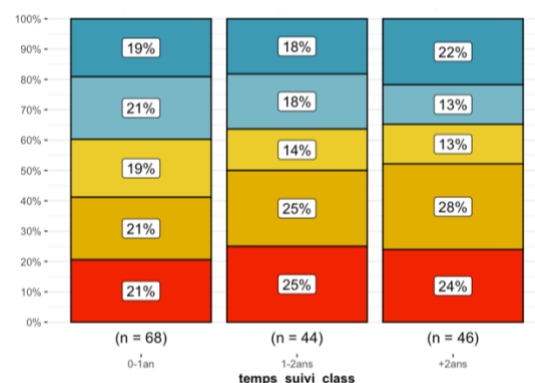


Figure 4: L'impact du temps de suivi des patients sur le niveau de diagnostic.

Impact du type d'EB sur l'évolution de la concentration de plomb chez les patients traités avec le même médicament

En considérant l'ensemble des patients soumis à la mesure des niveaux de plombs, la différence entre les semaines W0 à W6 est significative (test non-paramétrique de Friedman pour échantillons appariés, $p\text{-value} < 2.2e-16$).

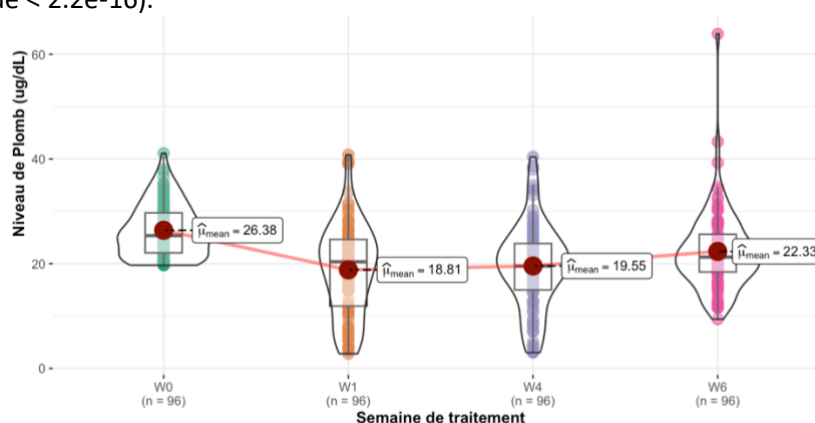


Figure 5: Évolution des niveaux de plomb pour tous les patients.

Lorsque l'on analyse côte-à-côte l'évolution des niveaux de plomb pour chaque population de patients, on constate qu'ils réagissent avec une dynamique différente. Les patients souffrant d'EB simple tendent à avoir un niveau de plomb qui chute fortement lors de la première semaine de traitement (W0) et qui remonte ensuite, tandis que les patients qui souffrent d'EB jonctionnelle ont un niveau de plomb qui descend progressivement au cours des 6 semaines de traitement (Fig.6).

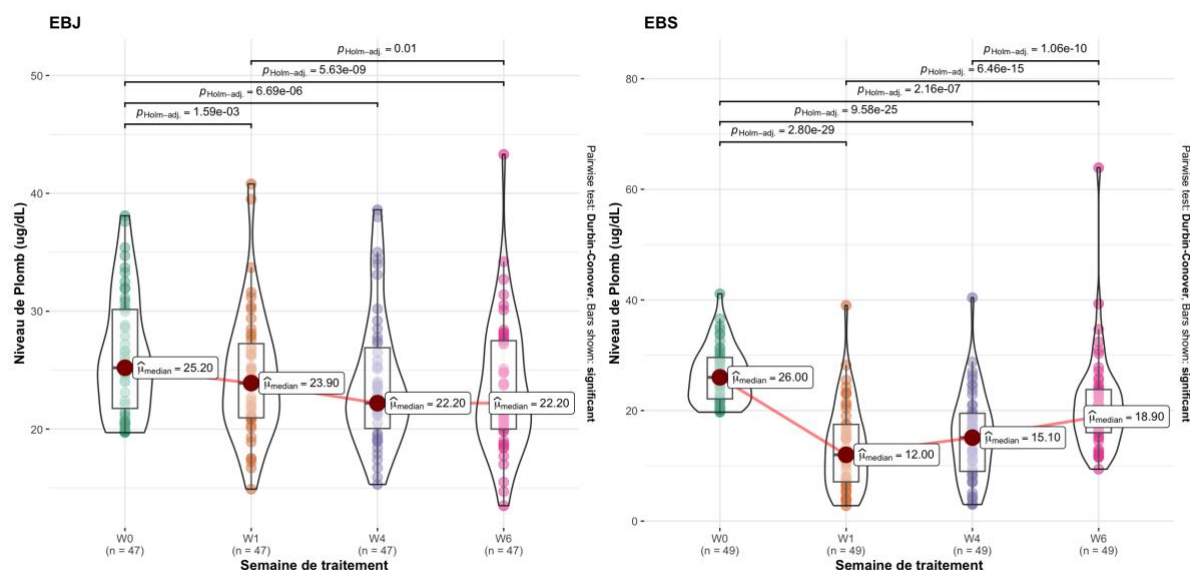


Figure 6: Évolution de la concentration en plomb dans le sang pour l'EB jonctionnelle et l'EB simple, respectivement.

Enfin, lorsque l'on compare les niveaux de plomb chaque semaine 2 à 2 pour l'EB jonctionnelle et l'EB simple, on s'aperçoit qu'ils sont significativement différents pour les semaines W1, W4 et W6. La plus forte différence observée est pour la semaine 1, au moment où la dynamique est vraiment différente entre les 2 formes d'EB.

Evolution du niveau de plomb au cours des semaines en fonction du type d'EB

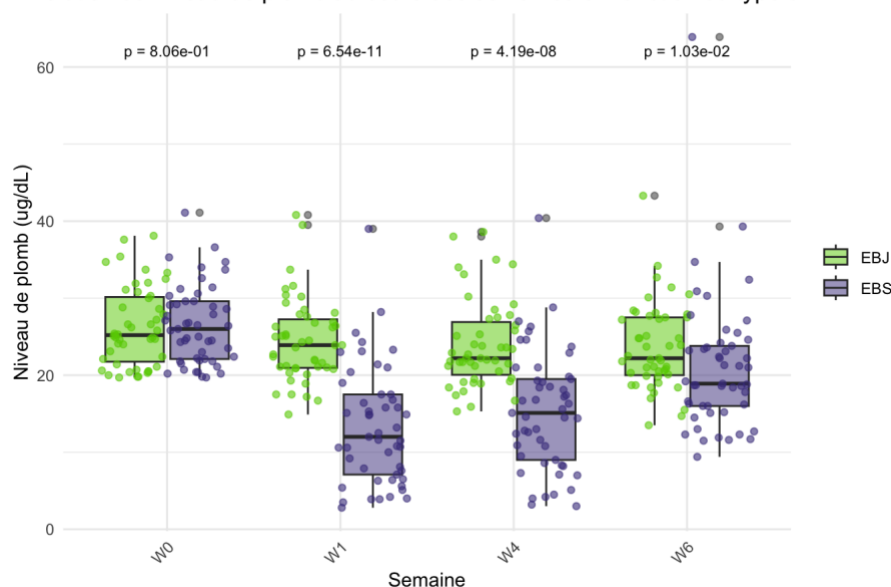


Figure 7: Comparaison des niveaux de plombs à chaque semaine en fonction du type d'EB. Les p-values ont été obtenues avec des tests de Wilcoxon (two-sided) pour comparer chaque diagnostic lors de chaque semaine deux à deux.

Il pourrait également être très intéressant d'étudier l'impact de l'âge (Annexe, Fig.11) sur l'effet du médicament et la fluctuation du niveau de plomb, et son interaction avec le type d'EB, car les deux sont certainement liés.

Notes sur le traitement des données

- Les variables à caractère personnel comme l'adresse et le numéro de maison n'ont surtout pas été utilisées, et écartées dès le début des analyses.
- La date de naissance, la date d'inclusion et la date de la dernière activité ont été recodées dans des variables catégorielles non identifiantes :
 - L'âge au moment de la dernière activité, en années.
 - Le temps de suivi entre la date d'inclusion et la date de la dernière activité, en jours.
 - Ces 2 variables quantitatives ont ensuite été découpées en classes pour lesquelles nous avons toujours plus de 10 individus par classe.
- Les données avec un identifiant dupliqué entre les différentes tables ont été écartées car elles ne permettaient pas de relier sans erreur un identifiant à un diagnostic. On ne peut pas les traiter comme 2 personnes et assigner un diagnostic au hasard à l'un des identifiants.
- La variable sur l'état du patient (vivant ou décédé) n'a pas été conservée car il y avait seulement 9 patients décédés et que son analyse n'apportait pas beaucoup aux conclusions.
- Les lignes avec des données manquantes (notamment après les jointures entre tables) ont systématiquement été écartées. On ne peut pas effectuer d'imputation de données manquantes dans ce contexte car nous souhaitons étudier une population réelle existante, de petite taille.
- Une donnée élevée, éventuellement un outlier, a été identifiée dans la semaine W6 des niveaux de plomb. En regardant les données pour cet individu, toutes ses mesures étaient particulièrement élevées. Il a été choisi de la conserver pour refléter la variation réelle de l'échantillon.
- Les tests de normalité et d'homogénéité de variance effectués avant de comparer les niveaux de plombs entre individus ont échoué. Ce n'est pas surprenant pour un échantillon biologique de cette taille, il faut donc appliquer des tests statistiques non paramétriques pour analyser ces données (i.e. : Friedman, Wilcoxon).
- Les deux datasets '_export.csv' contiennent des données sans paramètres identifiants qui pourrait être exportées, ré-utilisées ou transmises pour des analyses complémentaires.

Annexes

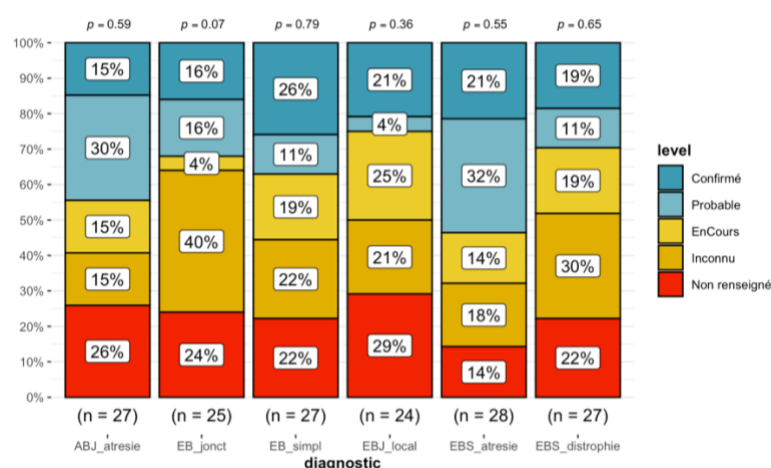


Figure 8: Niveau de diagnostic en fonction du sous-type d'EB.

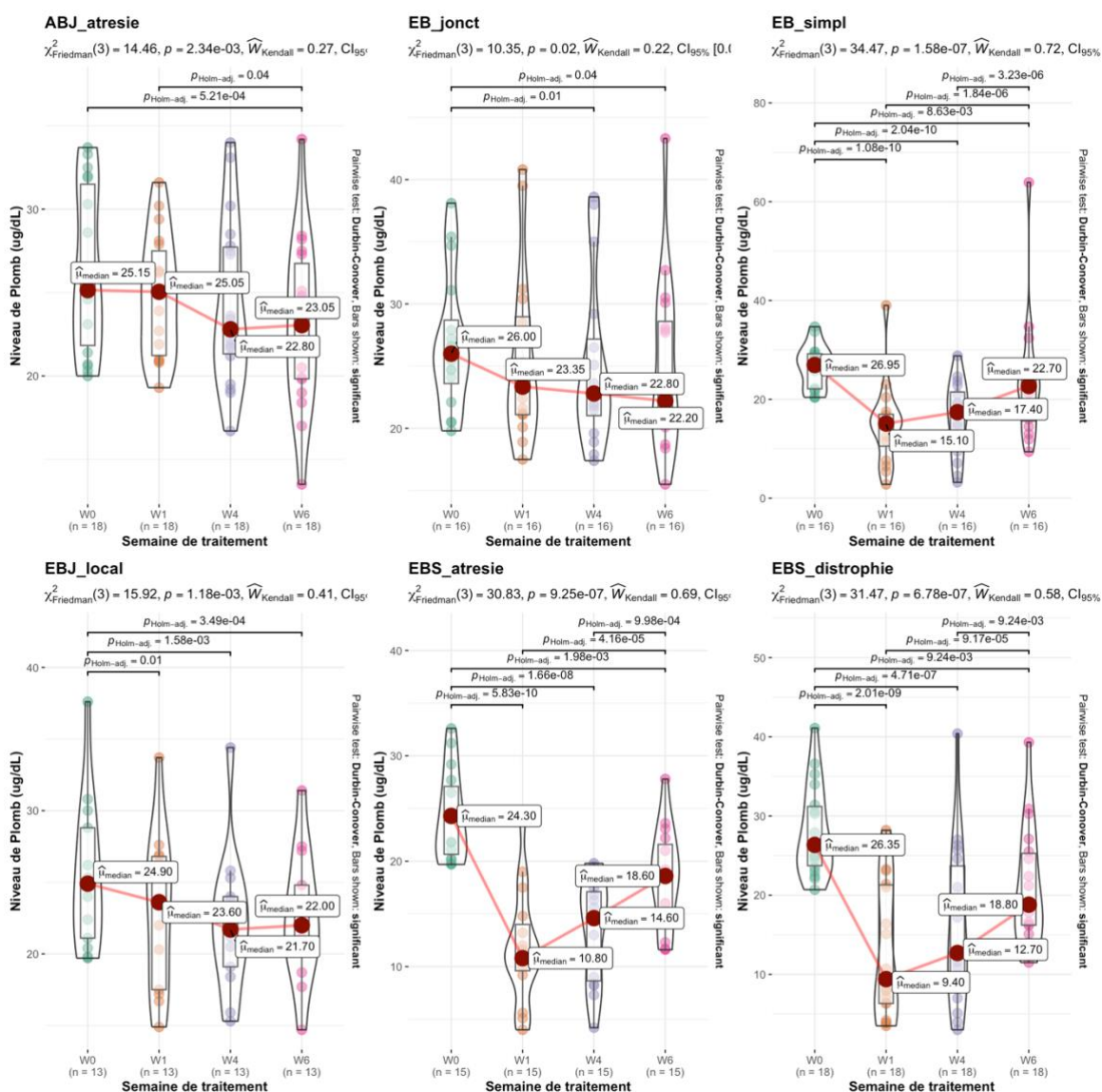


Figure 9: Evolution du niveau de plomb en fonction du sous-type d'EB. On observe globalement une tendance identique à celle pour le type d'EB général.

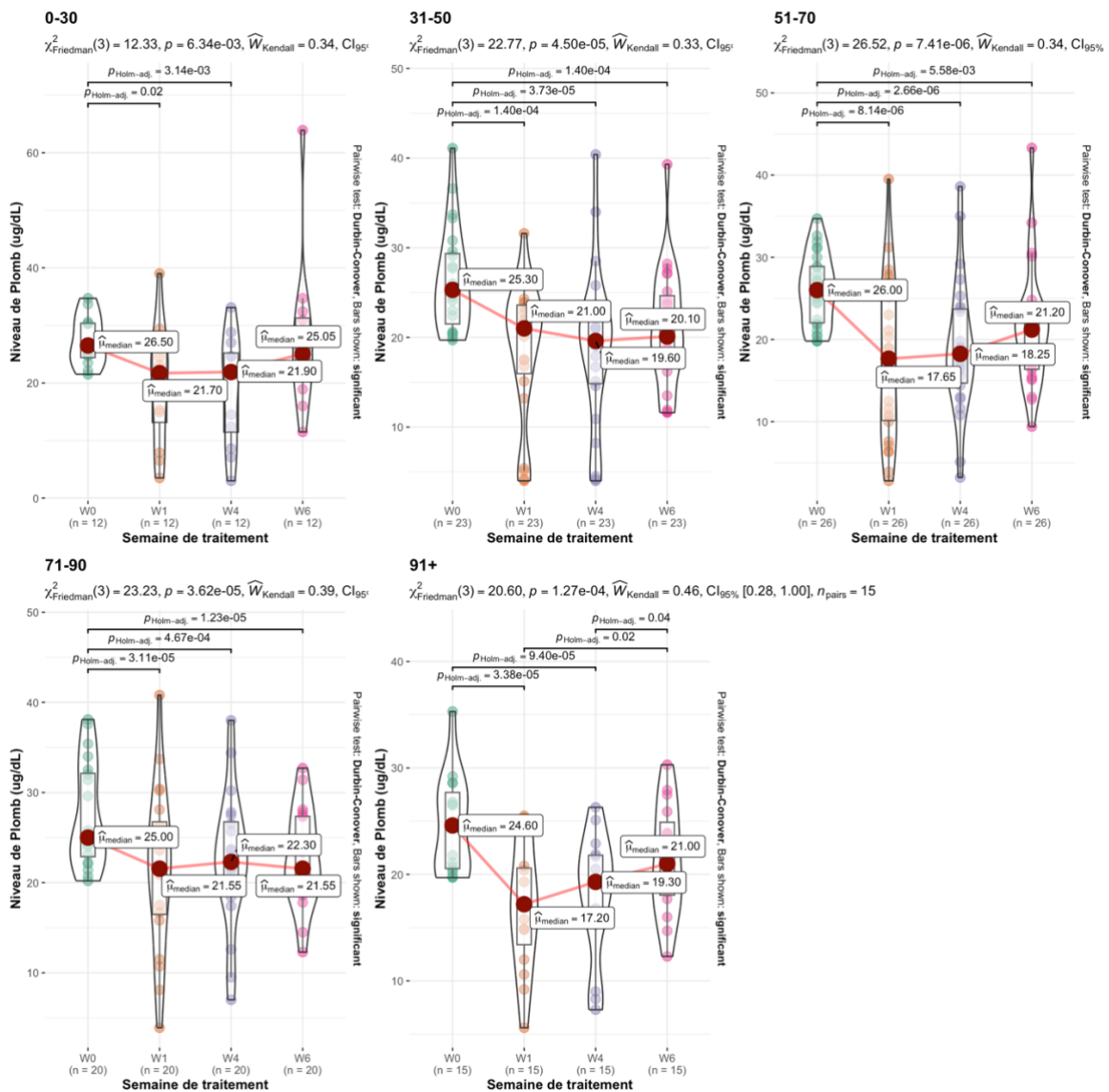


Figure 10: Évolution du niveau de plomb en fonction de la classe d'âge. Cela lisse les différences observées entre les types d'EB