



Projet adenome-pros

Guillaume Morin, Frédéric Saunier

Table des matières

1 INTRODUCTION

Cette étude porte sur 3 bases de données médicales VAPOR, RTUPB et VPPBS. Ces trois bases fournissent un ensemble de données pré et post opératoires pour un ensemble de patients utilisant l'un des trois traitements. Sachant que les données post opératoires sont fournies sous forme d'observations sur des intervalles de temps distincts.

TABLE 1 – GLOSSAIRE

Variable	Description
Age (ans)	Age du patient
Comorbidité CardioVx	Présence de maladies associées cardiaque ou vasculaire tel que l'hypertension artérielle
Durée traitement médical (mois)	N/A
Porteur de sonde	le patient a une sonde urinaire avant l'intervention
IPSS P.O	International prostatic symptom score PRE OPERATOIRE = plus il est élevé plus le patient est gêné
QoL P.O	Score de qualité de vie PRE OPERATOIRE = plus il est élevé plus le patient est insatisfait
Qmax P.O (ml/s)	Débit maximal urinaire PRE OPERATOIRE = plus il est élevé, plus la miction est de bonne qualité
PSA (ng/ml)	N/A
Volume prostatique (ml)	N/A
RPM	Residu post mictionnel = quantité d'urine retrouvée dans la vessie après une miction, à l'état normal elle est de 0
Indication	N/A
Anesthésie	N/A
Evenement H.D	Evénement hémodynamique pendant l'intervention = perturbation de la tension artérielle durant l'intervention
Transfusion PerO	Si oui ou non le patient a eu une transfusion pendant l'intervention
Temps OP	Temps opératoire
Volume résequé (ml)	N/A
Délai ablation (jours)I)	Délai d'ablation de la sonde urinaire après l'intervention
caillotage	N/A
reprise au bloc	N/A

2 ANALYSE DESCRIPTIVE

2.1 VUE GÉNÉRALE

RTUPB, VPPBS et VAPOR sont trois bases d'observations contenant respectivement 36, 32 et 40 observations (patients) avec une répartition en âge suivante :

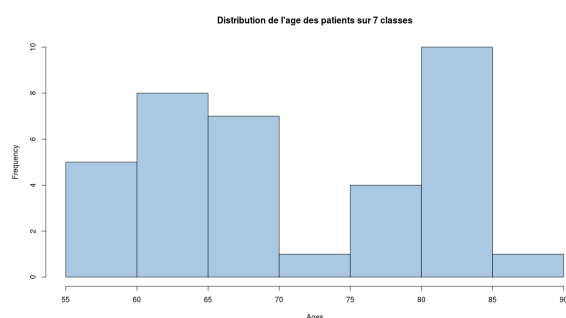


FIGURE 1 – RTUPB

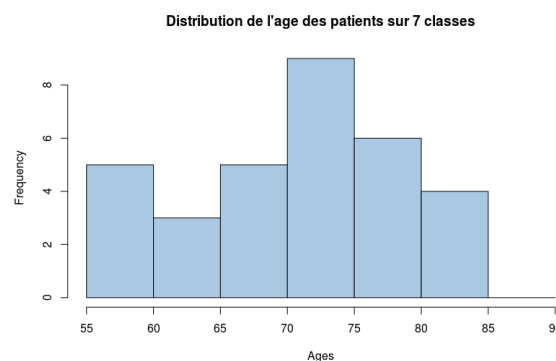


FIGURE 2 – VPPBS

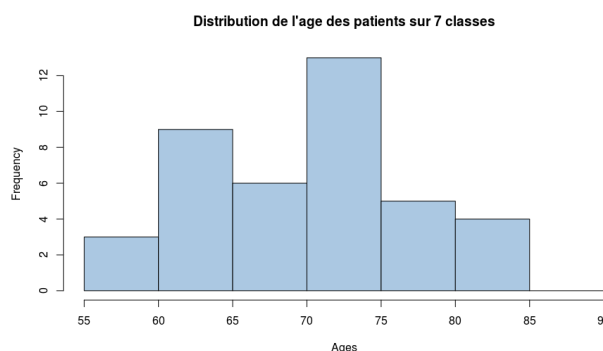


FIGURE 3 – VAPOR

2.2 CORRÉLATION VARIABLES PRÉ-OPÉRATOIRES

Ici nous souhaitons mettre en évidence les corrélations possibles. Nous excluons ici l'étude sur les tableaux post-opératoires des techniques RTUPB VPPBS et VAPOR contenant aussi les variables IPSS Qol et Qmax. De même, certaines dimensions sont invariantes pour une technique donnée, nous avons supprimé celles-ci lors de la création de la matrice de corrélation et de son corrélogramme associé.

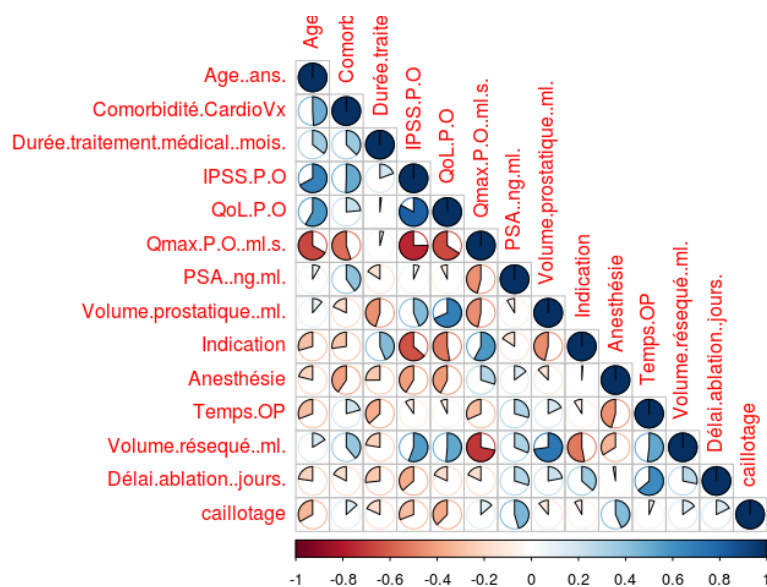


FIGURE 4 – MATRICE CORRÉLATION RTUBP

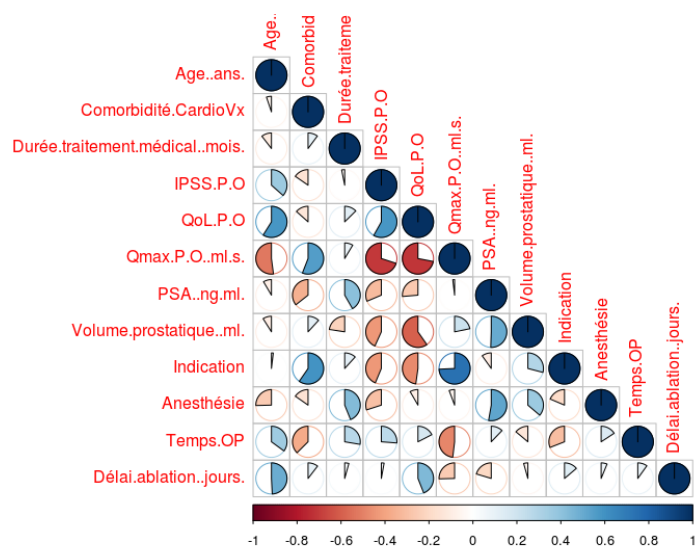


FIGURE 5 – MATRICE CORRÉLATION VPPBS

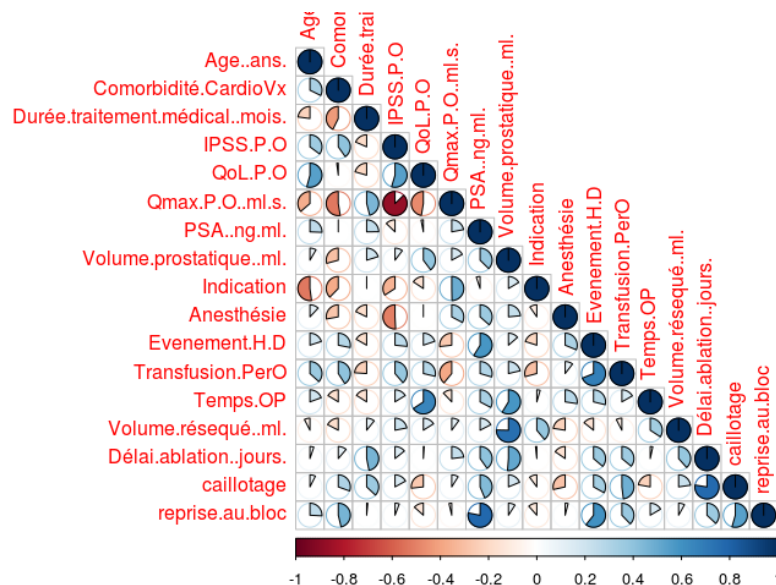


FIGURE 6 – MATRICE CORRÉLATION VAPOR

Les variables *IPSS P.O* et *QoL P.O* semblent avoir une corrélation qui peut sembler logique à la connaissance du fait qu'elles représentent pour l'une un indicateur de gêne et pour l'autre un indicateur une qualité de vie post opératoire, même si c'est nettement plus marqué dans le cas du panel des patients RTUPB. De même pour les variables *Volume prostatique* et *Volume résequé*. Aussi nous avons une corrélation **négative** intéressante entre le *IPSS P.O* et le *QMAX PO (ml/s)* (plus le patient a un QMax élevé moins il semble gêné alors que IPSS croît avec la gêne).

2.3 DISTRIBUTION/ÉVOLUTION DES DONNÉES POST-OPÉRATOIRES

Dans l'analyse suivante nous souhaitons observer la distribution des variables IPSS, QoL et Qmax sur les différentes itérations temporelles fournies. L'objectif est de voir (pour l'ensemble des patients utilisant l'une des trois techniques) comment cette distribution évolue.

2.3.1 IPSS SUR 18 MOIS

RTUPB est une table composée de 36 patients.

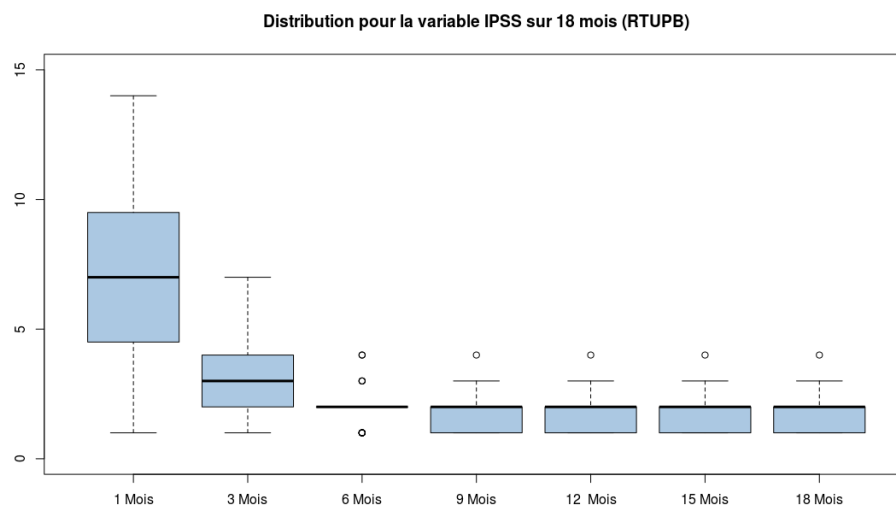


FIGURE 7 – RTUPB / IPSS SUR 18 MOIS

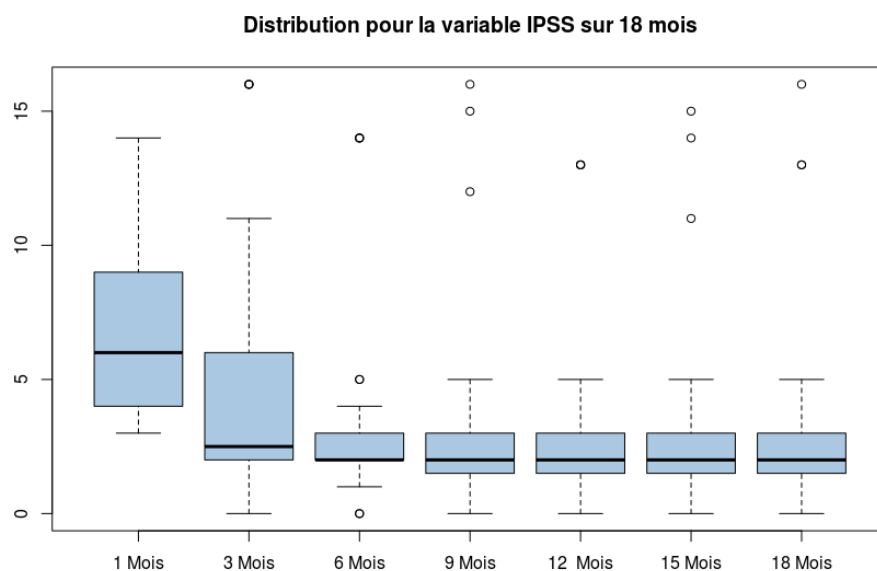


FIGURE 8 – VPPBS/IPSS SUR 18 MOIS

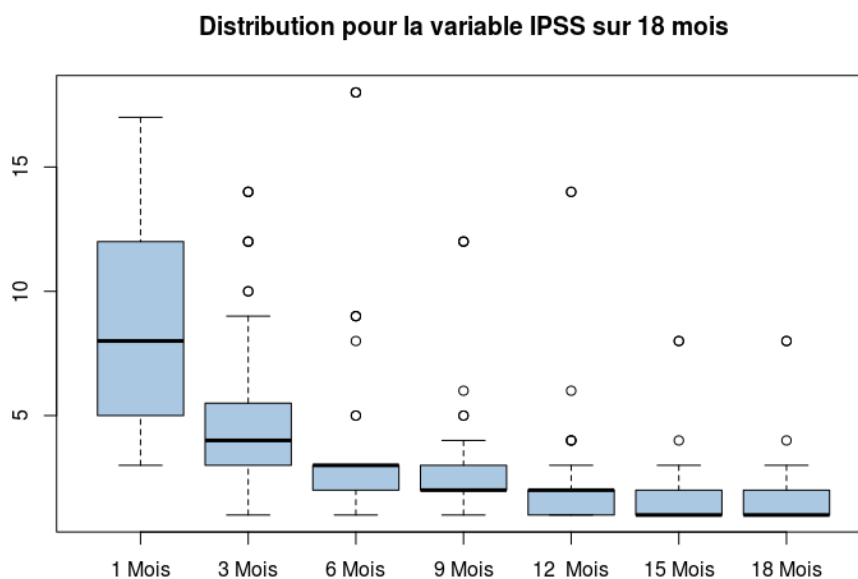


FIGURE 9 – VAPOR/IPSS

Pour IPSS, on observe pour les trois techniques une décroissance de la valeur médiane dès le troisième mois. Dans le cadre de VAPOR et VPPBS, certains « outliers » sont présents (et ce sur l'ensemble des itérations dans le cadre de VPPBS ces « outliers » ont encore des valeurs très fortes même à partir du 18^{ième} mois. RTUPB semble avoir une variance homogène à partir du 9^{ième} mois. Si VAPOR laisse apparaître des « outliers », la valeur médiane termine plus bas que les deux autres techniques.

2.3.2 QOL SUR 18 MOIS

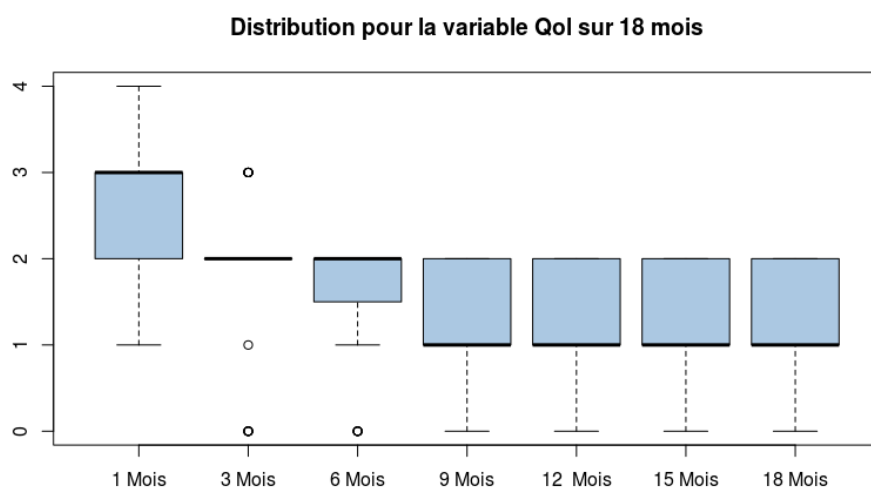
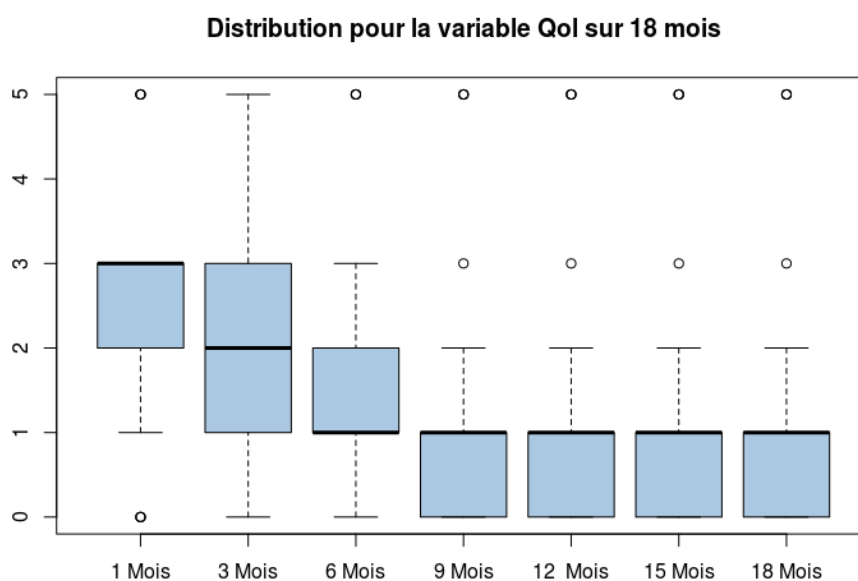
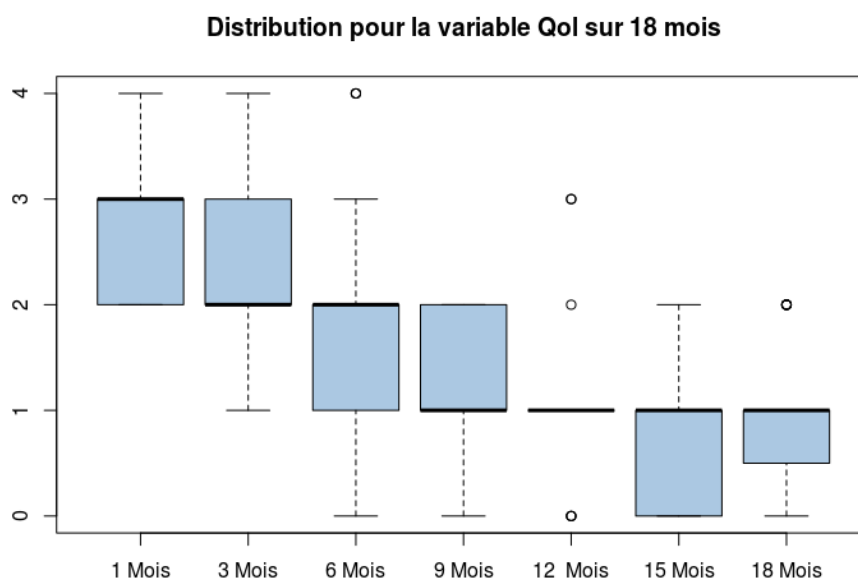


FIGURE 10 – RTUPB / QOL SUR 18 MOIS

**FIGURE 11 – VPPBS / QOL SUR 18 MOIS****FIGURE 12 – VAPOR / QOL SUR 18 MOIS**

Les valeurs de QoL pour RTUPB et VPPBS semblent être assez stables après le 9^{ième} mois, dans le cas de VAPOR on peut constater une variabilité plus élevée.

2.3.3 QMAX SUR 18 MOIS

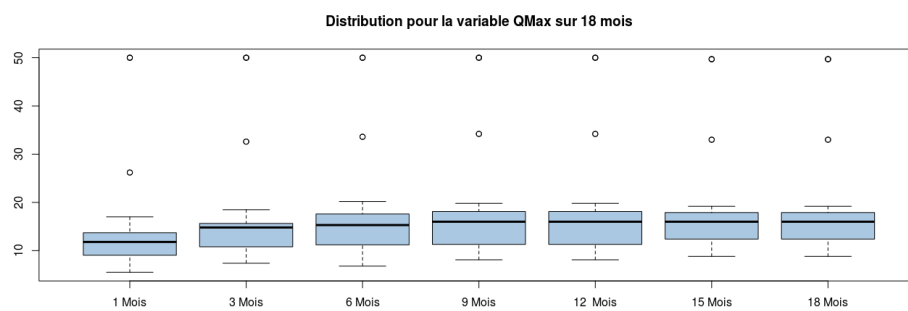


FIGURE 13 – RTUPB / QMAX SUR 18 MOIS

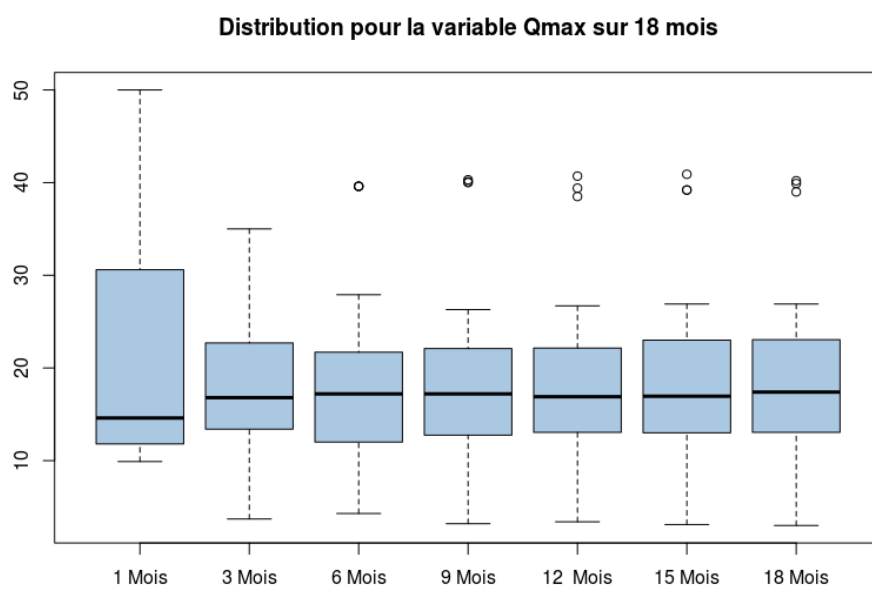


FIGURE 14 – VPPBS/QMAX SUR 18 MOIS

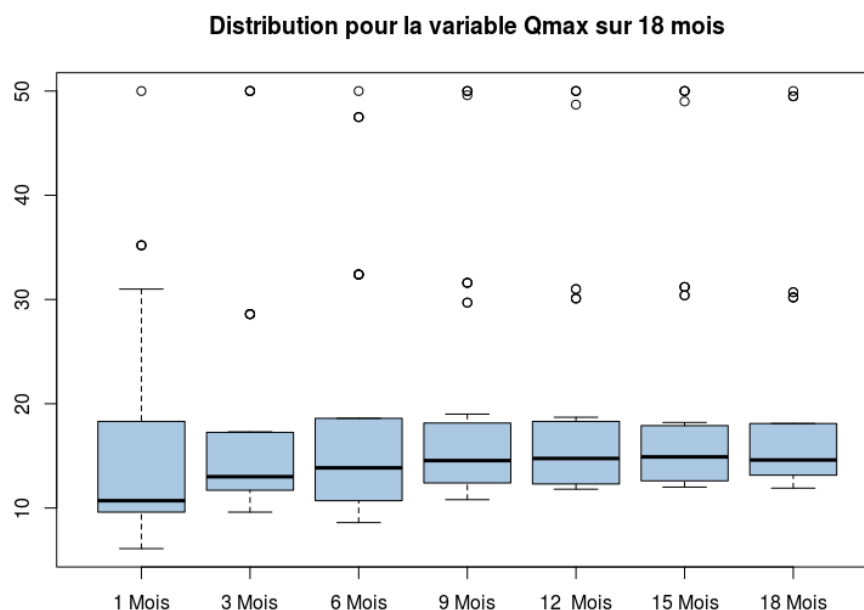


FIGURE 15 – VAPOR/QMAX

RTUPB semble être le plus constant, VPPBS permet d'obtenir de meilleurs résultats sur le Qmax (augmentation) avec des cas extrêmes plus élevés. VAPOR obtient des résultats médians proches de RTUPB mais avec des cas extrêmes plus élevés comparativement aux trois techniques.

3 CLASSIFICATION PROFILS PRÉ-OPÉRATOIRES

Dans le cadre de la classification nous avons observé quelques doublons, nous avons choisi de les supprimer du moins dans cette première phase.

3.1 CAH / PAM RTUPB PRÉ-OPÉRATOIRES

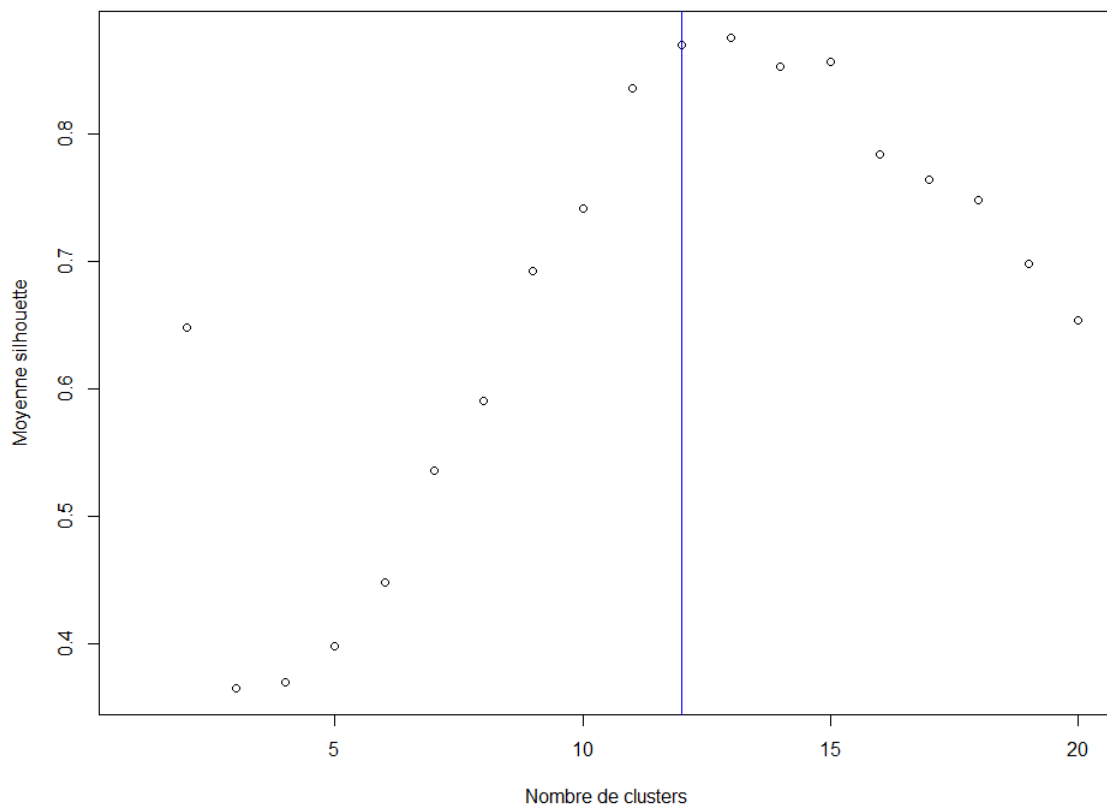


FIGURE 16 – RTUPB MAXIMISE NB CLUSTER / BONNE CLASSIFICATION

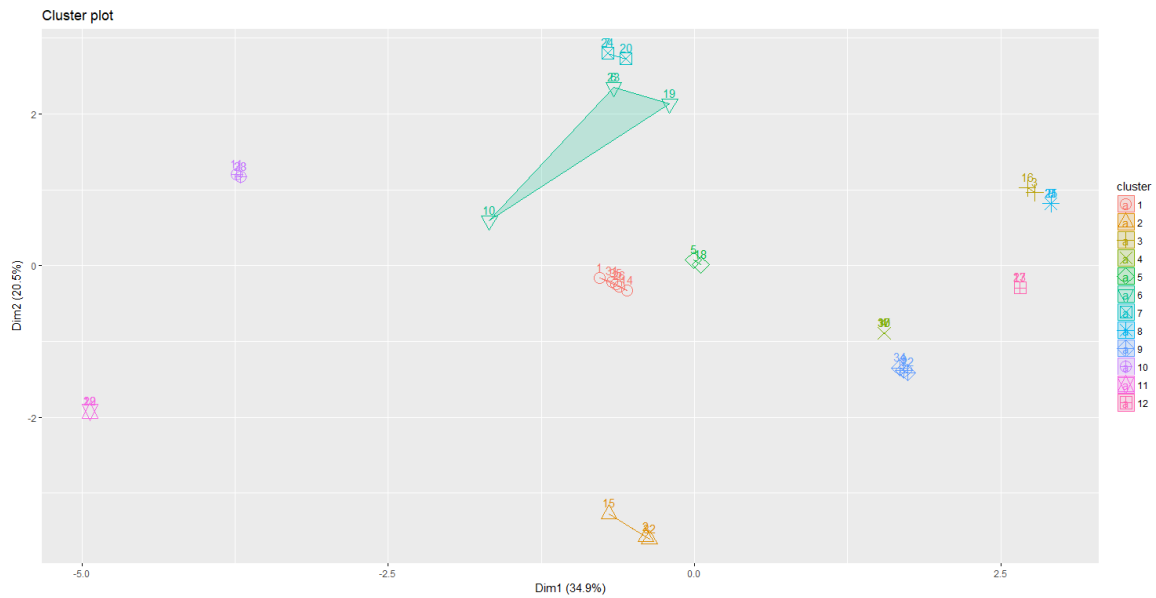


FIGURE 17 – RTUPB NUAGE DE POINTS / CLUSTERS PAM K=12

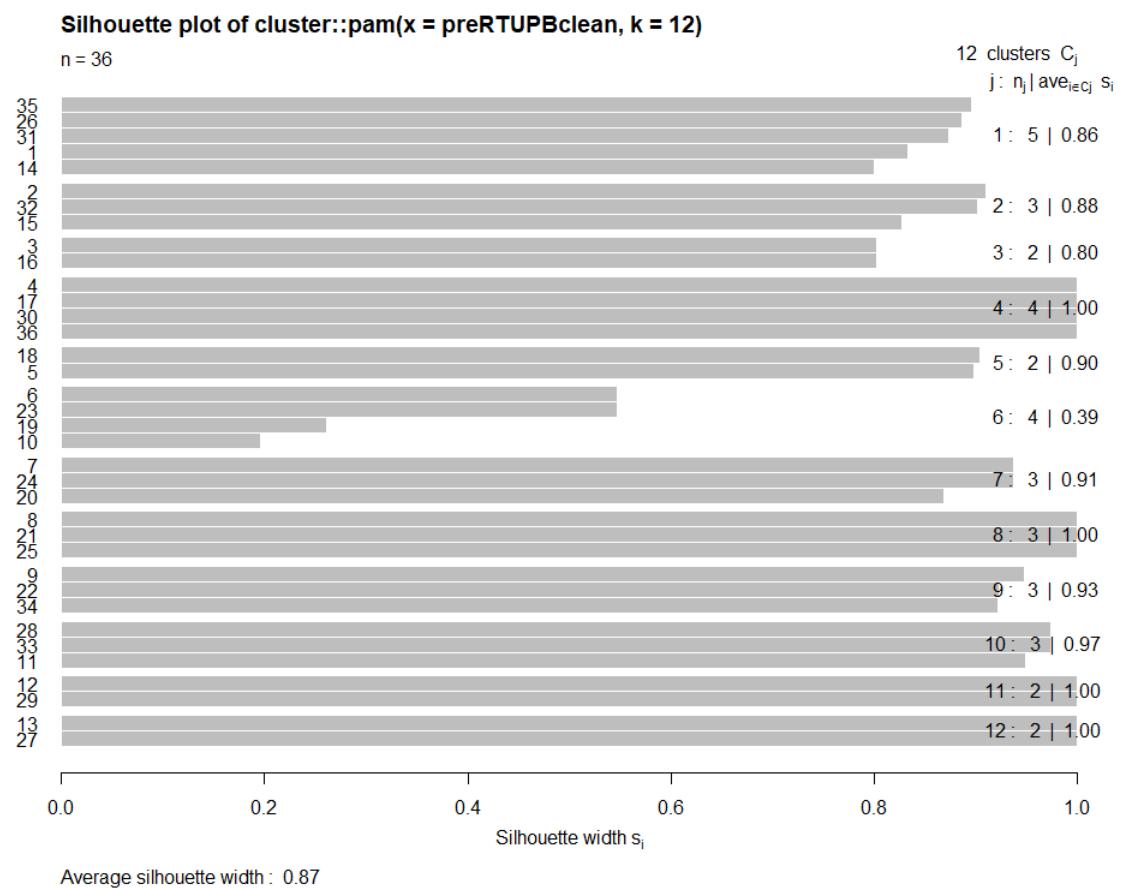


FIGURE 18 – RTUPB SILHOUETTE / CLUSTERS PAM K=12

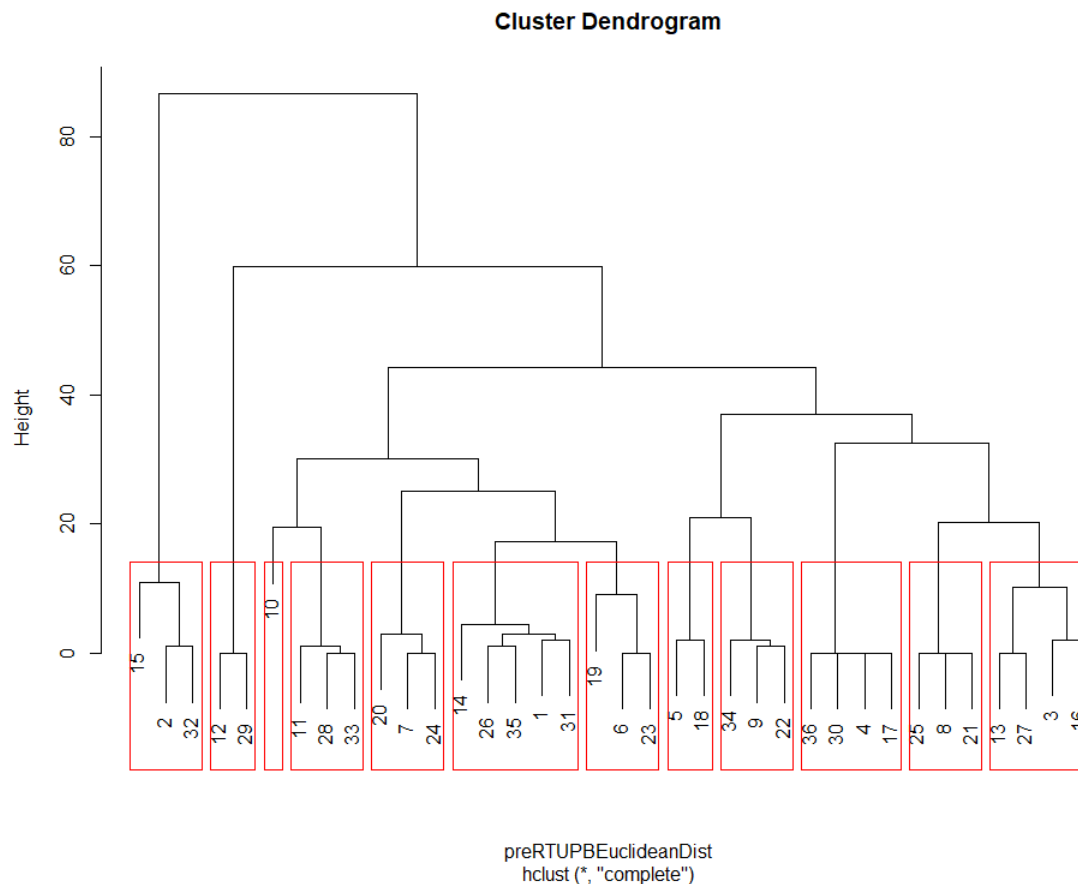


FIGURE 19 – RTUPB CAH SÉPARATION EN K=12

Lors de notre analyse nous avons observé qu'il existait un ensemble d'individus similaires (ensemble des valeurs identiques au relevé prés) ce qui s'observe dans la construction de certains clusters avec une valeur de qualité intra-cluster très forte en forgeant de petits clusters. Nous reviendrons sur ce point dans le rapprochement de ces profils avec les profils de guérison ultérieurement. Nous avons estimé à 12 le nombre de clusters suivant la variation de la qualité globale du cluster réalisé à partir de PAM.

3.2 CAH / PAM VPPBS PRÉ-OPÉRATOIRES

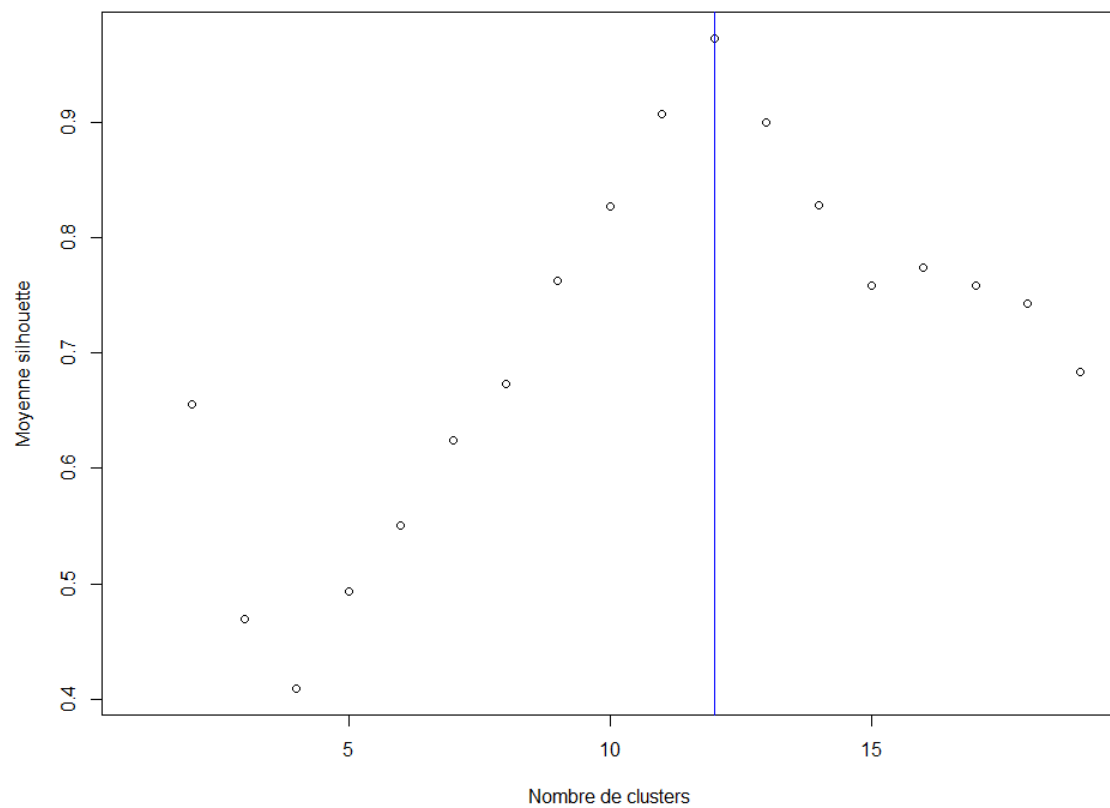


FIGURE 20 – VPPBS MAXIMISE NB CLUSTER / BONNE CLASSIFICATION

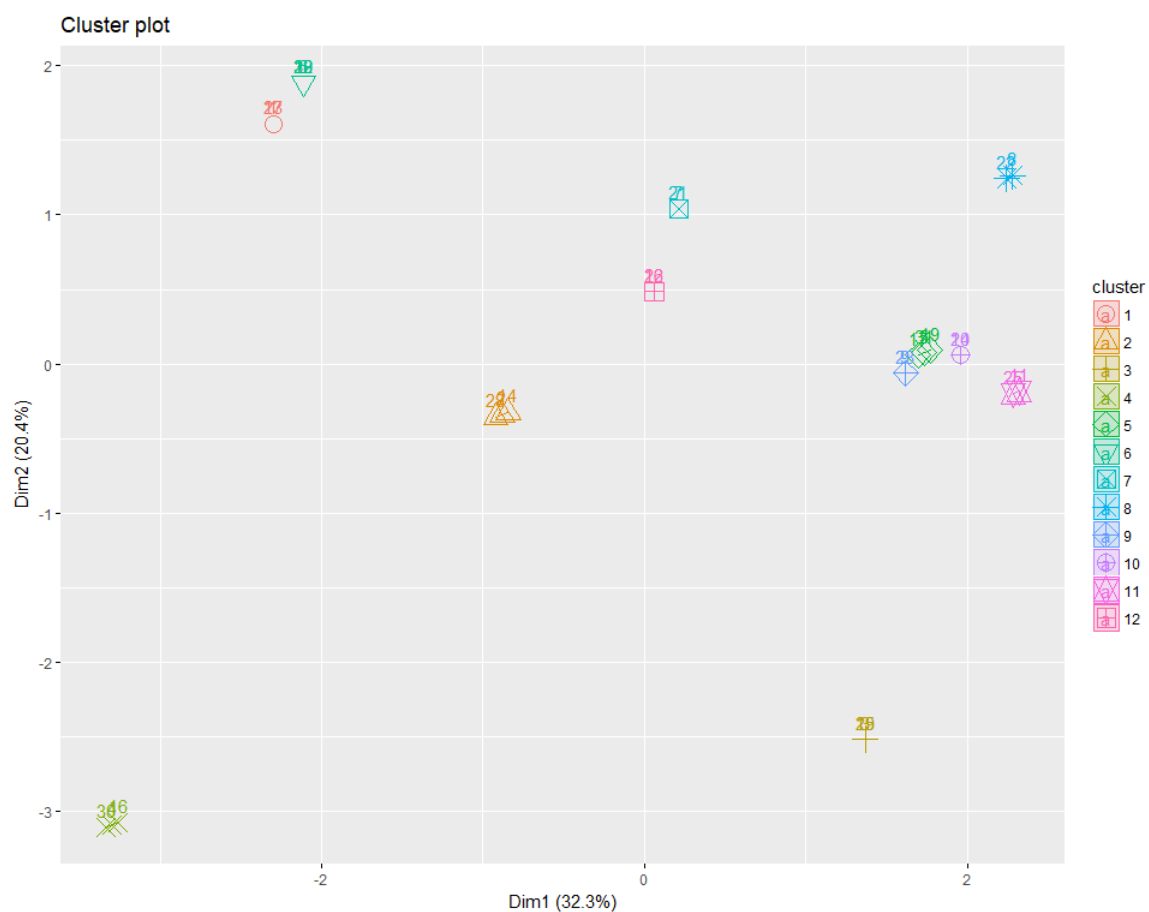


FIGURE 21 – VPPBS NUAGE DE POINTS / CLUSTERS PAM K=12

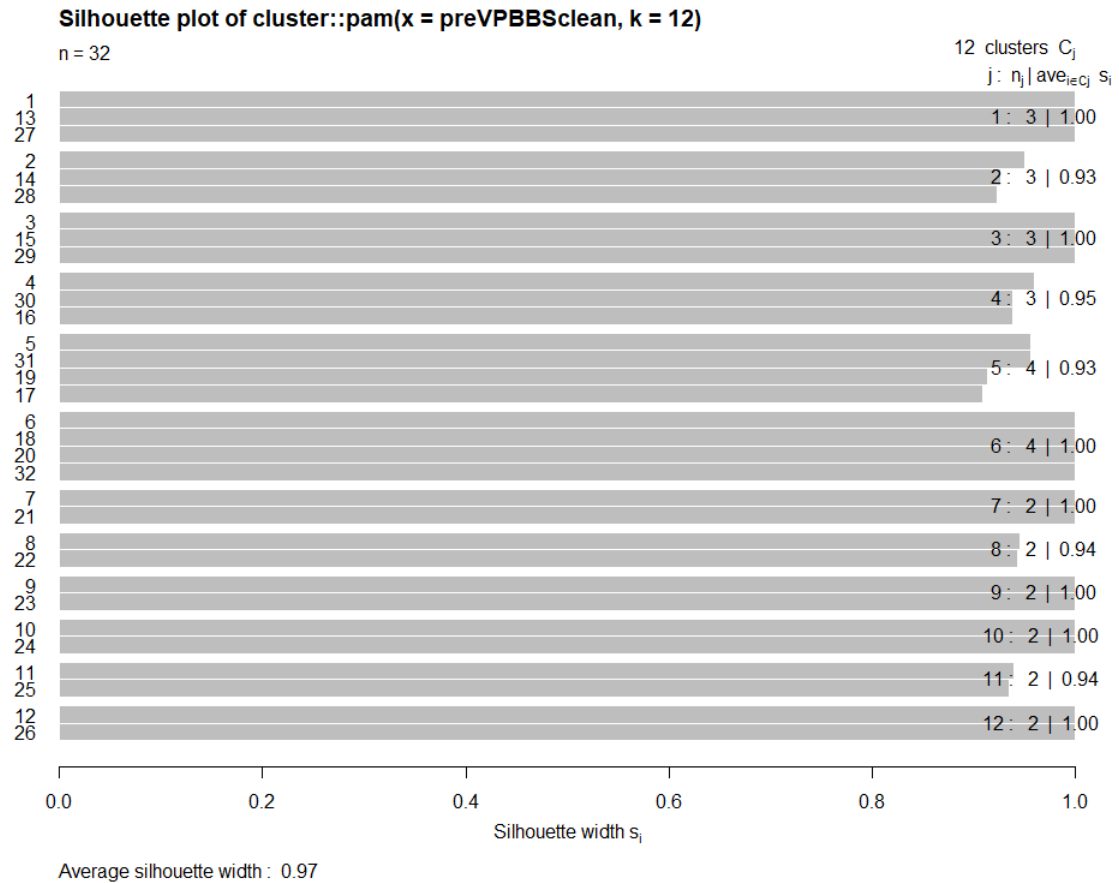


FIGURE 22 – VPPBS SILHOUETTE/ CLUSTERS PAM K=12

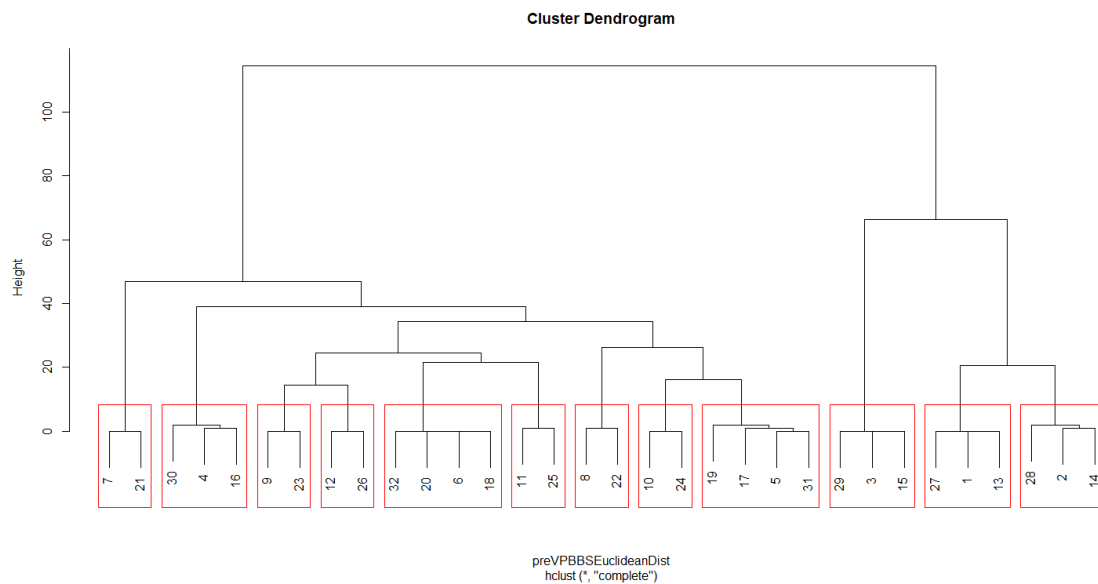


FIGURE 23 – VPPBS CAH SÉPARATION EN K=12

3.3 COMPARAISON QMAX

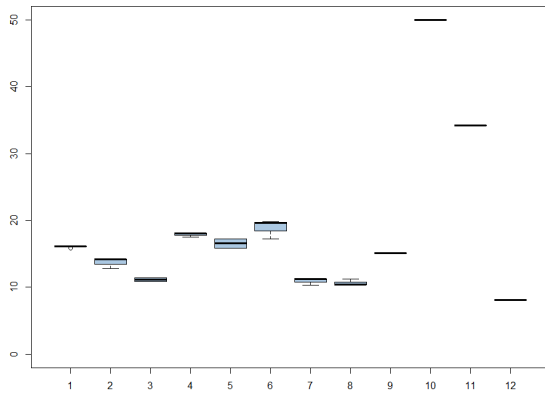


FIGURE 24 – DISTRIBUTION QMAX À 12 MOIS RTUPB

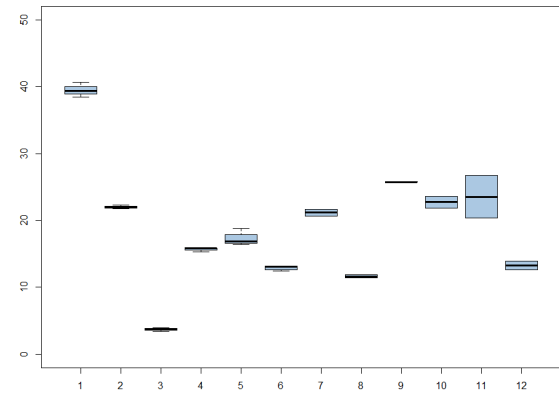


FIGURE 25 – DISTRIBUTION QMAX À 12 MOIS VPPBS

4 CLASSIFICATION PROFILS POST-OPÉRATOIRES

4.1 CAH / PAM RTUPB POST-OPÉRATOIRES

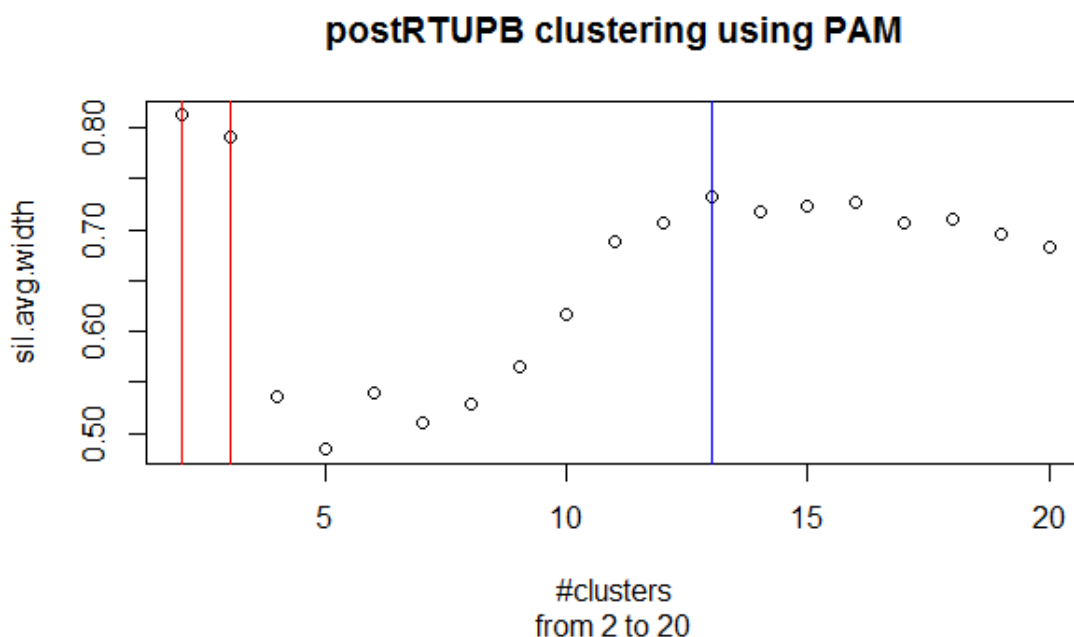


FIGURE 26 – MAXIMISATION DE LA SILHOUETTE MOYENNE

En utilisant PAM sur les données post-opératoires, la courbe des silhouettes moyennes nous indique une valeur maximale pour un nombre de classes à $k = 2$ (Cf. figure ??).

Cependant, si nous regardons le détail des classes et de leurs silhouettes pour $k = 2$ (Cf. figure ??), nous avons une première classe incluant la quasi totalité des patients dont 2 qui ont un profil post-opératoire non relié (valeurs de silhouette négatives), et une deuxième classe

ne contenant que trois patients (11, 28 et 33) avec une valeur de silhouette à 1, indiquant des données répliquées exactement et donc une classe triviale. Après vérification, nous constatons que ces patients ont des données post-opératoires identiques, mais ne sont pas pour autant des doublons (données pré-opératoires différentes). Nous poursuivons la recherche jusqu'au maximum suivant qui se trouve à $k = 3$ (Cf. figure ??).

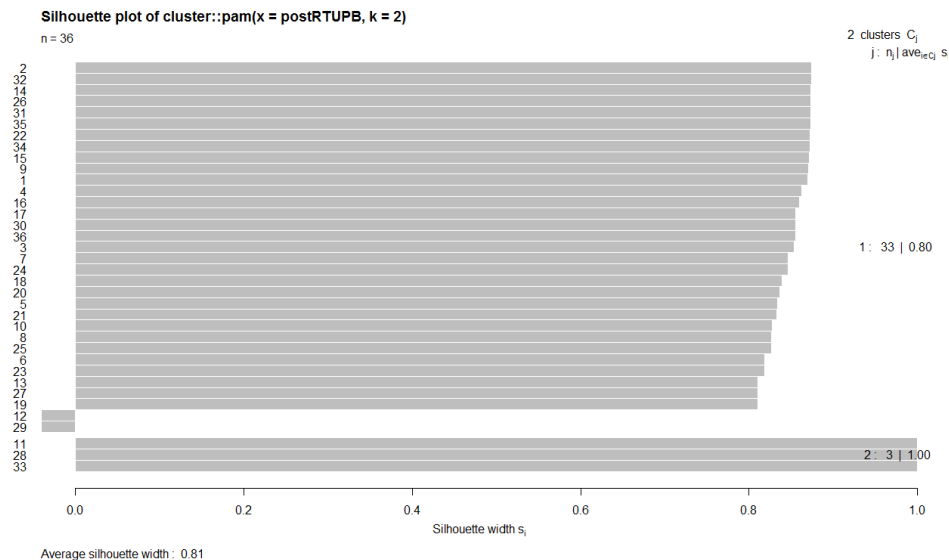


FIGURE 27 – SILHOUETTE / CLASSE (K = 2)

En examinant, pour $k = 3$ (Cf. figure ??), la répartition des classes et leurs silhouettes, nous faisons le même constat que précédemment, et une nouvelle classe triviale (patients 12 et 29) est apparue pour des patients distincts mais dont les données post-opératoires sont identiques. Nous poursuivons encore la recherche jusqu'au maximum suivant qui se trouve à $k = 13$ (Cf. figure ??).

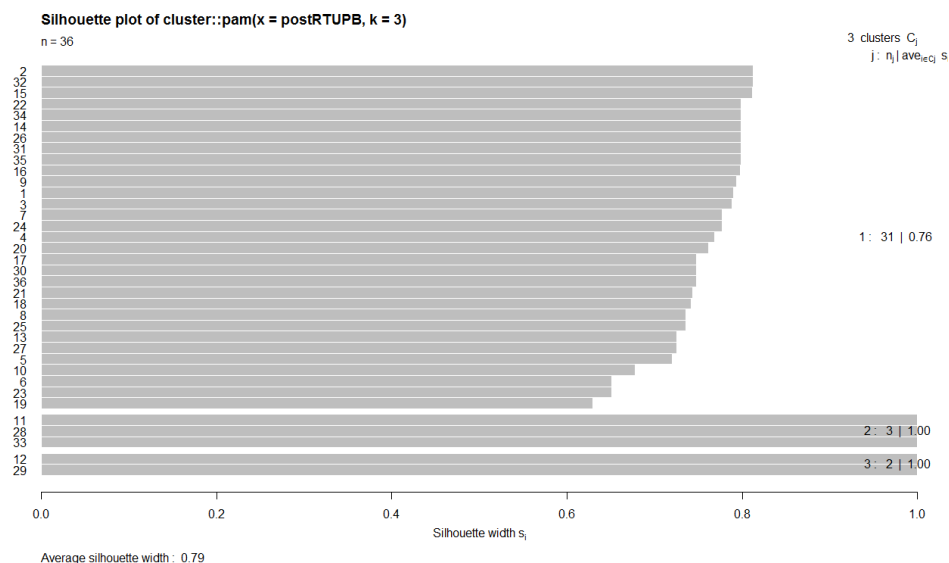


FIGURE 28 – SILHOUETTE / CLASSE (K = 3)

Le partitionnement à 13 classes permet d'identifier 10 classes de profils post-opératoires bien, voire très bien classés : 3 dont la silhouette vaut 1, 5 dont la silhouette est comprise entre 0.75 et 1, et 2 dont la silhouette est aux alentours de 0.6). Ce partitionnement fait cependant apparaître 2 classes de profils post-opératoires pas très bien classés (silhouette < 0.5) et une classe singleton (silhouette = 0).

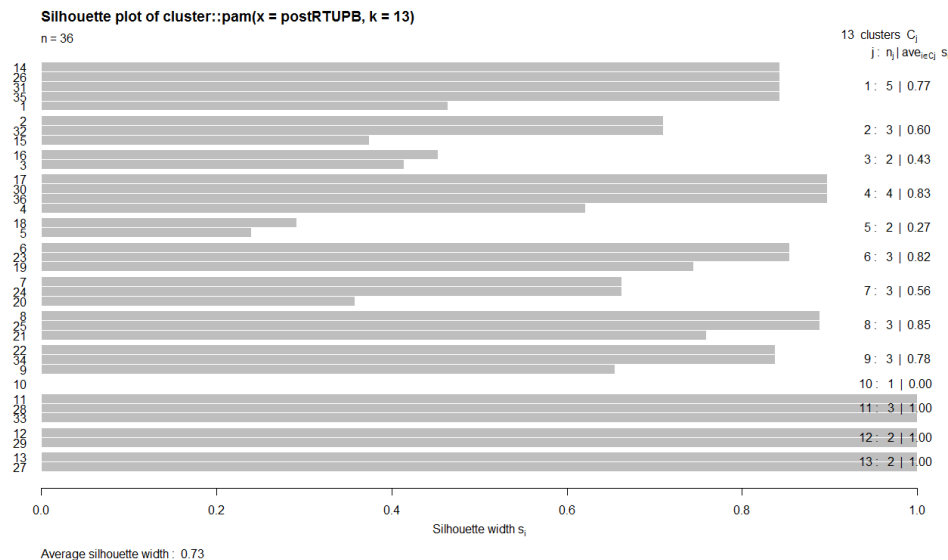


FIGURE 29 – SILHOUETTE / CLASSE (K = 13)

La classification hiérarchique favoriserait plutôt un niveau de coupure de l'arbre correspondant à un partitionnement à 4 classes représenté en bleu sur la figure ???. Un partitionnement avec des classes plus proches nous conduit à couper l'arbre au niveau suivant, ce qui nous donne alors 13 classes. Ce partitionnement est représenté en bleu sur la figure ???. Ce partitionnement à 13 classes correspond à celui obtenu par la méthode PAM, nous allons le retenir pour étudier dans la suite, les profils post-opératoires des patients médoïdes, et si cela se justifie après cela, nous refactoriserons en seulement 4 classes.

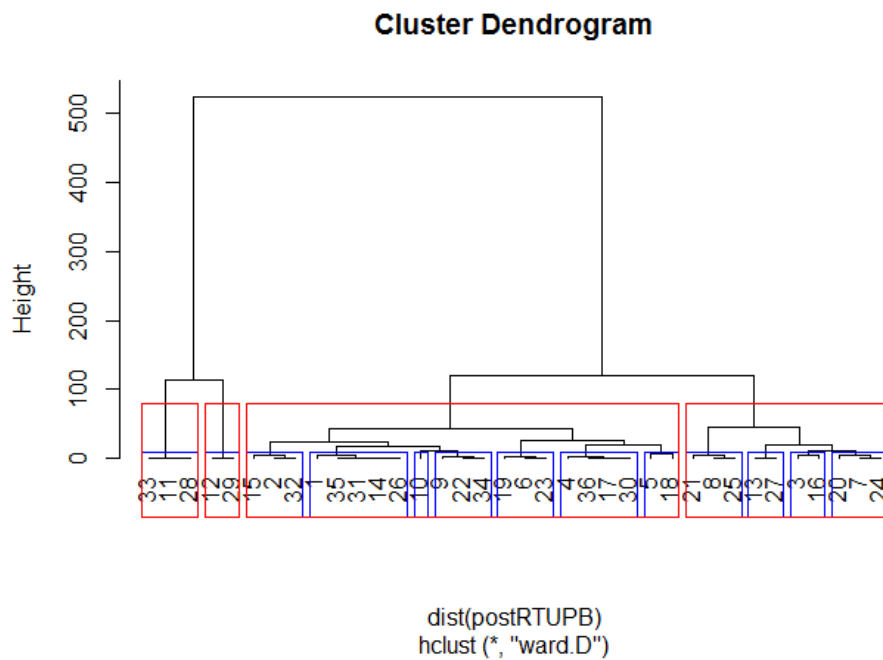


FIGURE 30 – SILHOUETTE / CLASSE (K = 13)

4.2 CAH / PAM VPPBS POST-OPÉRATOIRES

Pour rechercher le nombre optimal de classes, nous avons évalué, figure ?? la silhouette moyenne des classes pour un partitionnement de 2 à 20 classes. Il en ressort un pic net (en rouge) à 4 classes, mais la valeur maximale de la silhouette moyenne est atteinte pour k = 12 classes.

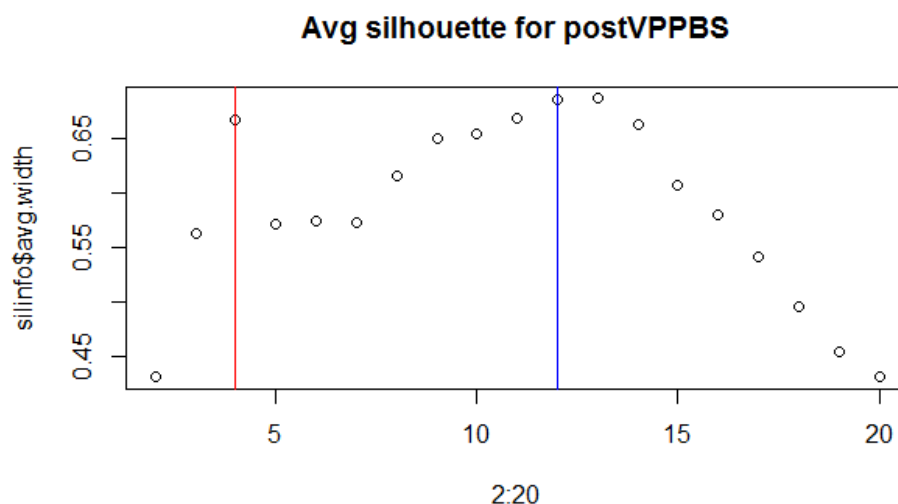


FIGURE 31 – RECHERCHE DU MEILLEUR NOMBRE DE CLASSES

Jetons un bref coup d'oeil sur les silhouettes de classes pour le partitionnement à 4 classes (Cf. figure ??). Nous pouvons y observer 3 classes très solides (valeurs de silhouette entre 0.9 et

1), non triviales (contrairement à certaines classes observées pour les données post-opératoires RTUPB), et chacune de ces classes compte 3 ou quatre patients; la quatrième classe rassemble tous les autres profils (22 patients sur 32), avec une silhouette de classe assez faible (0.55). Les 3 premières classes observées ici se détachent nettement et nous les retrouverons donc dans le partitionnement à $k = 12$ classes qui, compte tenu de la silhouette moyenne maximale, devrait faire apparaître de nouvelles classes parmi les 22 profils restant. Les silhouettes correspondant à ce partitionnement à 12 classes font ressortir, figure ?? :

- 6 classes de profils fortement corrélés (silhouette > 0.7)
- seulement 2 classes de profils peu ou pas corrélés (silhouette < 0.5)

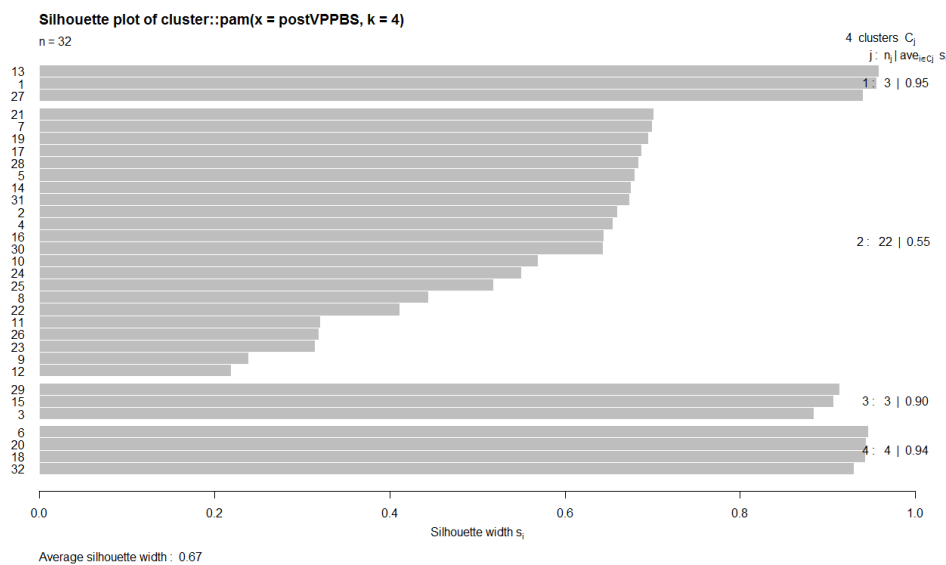


FIGURE 32 – SILHOUETTE / CLASSE (K = 4)

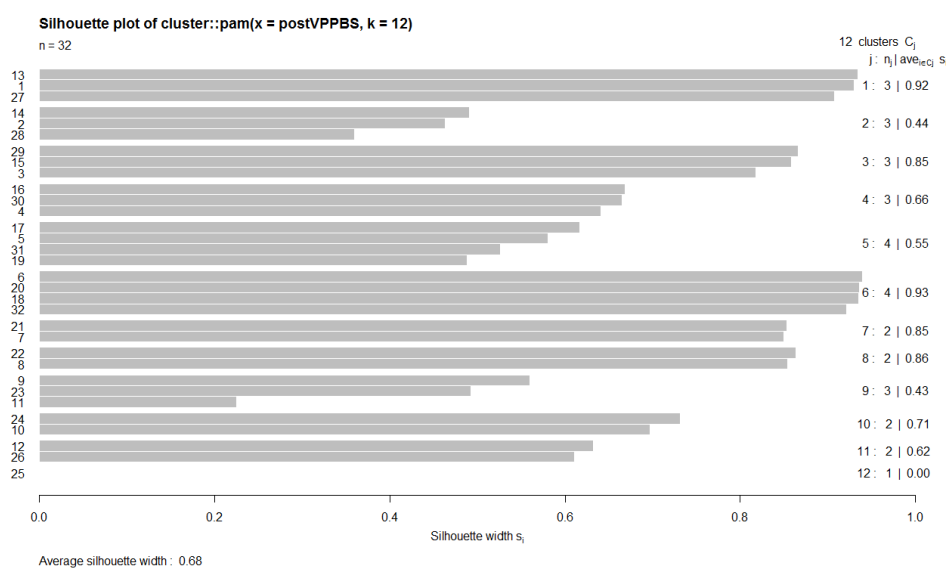


FIGURE 33 – SILHOUETTE / CLASSE (K = 12)

La classification hiérarchique des données post-opératoires VPPBS produit le dendrogramme

présenté figure ?? La coupure de l'arbre peut s'effectuer soit à 5 classes (représentées en rouge), soit à 12 classes (représentées en bleu). Le partitionnement à 12 classes correspondant à celui obtenu par la méthode, nous le retiendrons pour étudier les profils post-opératoires. Au besoin, si cela se justifie par la suite, nous factoriserons pour revenir à un partitionnement à 5 classes.

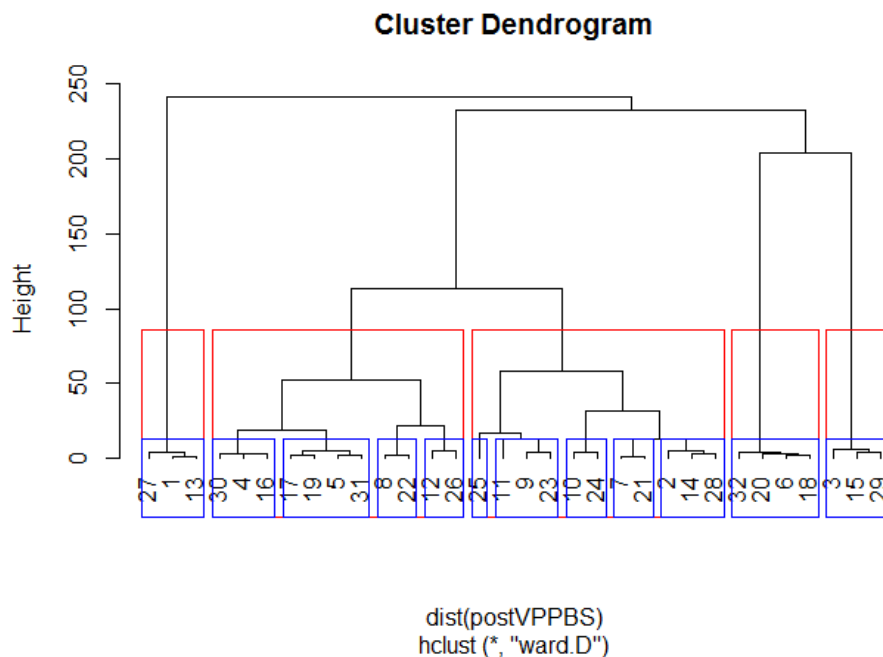


FIGURE 34 – VPPBS : CLASSIFICATION HIÉRARCHIQUE

4.3 PROFILS PRÉ-OPÉRATOIRES

4.4 EXTRACTION DES PROFILS POUR RTUPBS

4.4.1 QMAX SUR 12 MOIS

4.5 EXTRACTION DES PROFILS POUR RTUPBS

4.5.1 QMAX SUR 12 MOIS

4.6 EXTRACTION DES PROFILS POUR RTUPBS

4.6.1 CONCLUSION

4.7 PROFILS POST-OPÉRATOIRES

4.7.1 RTUPB

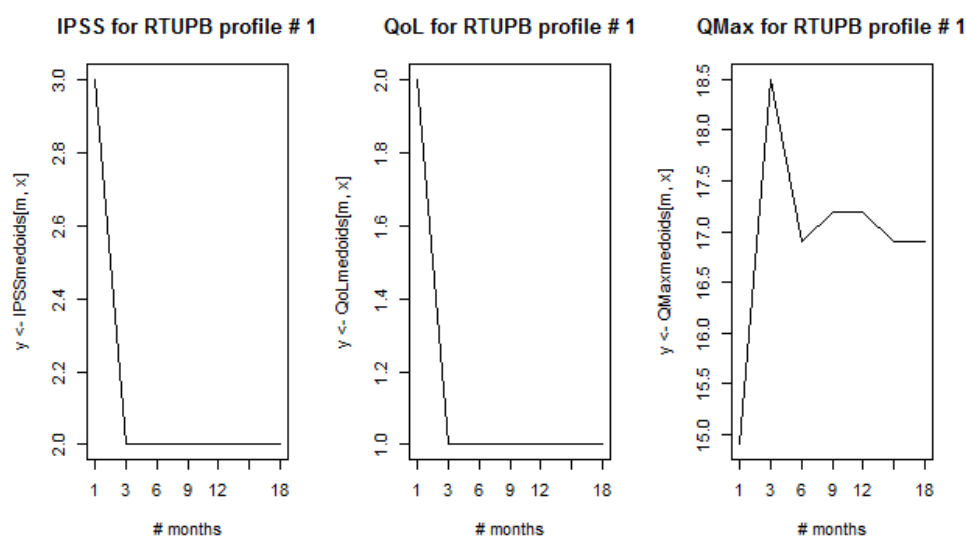


FIGURE 35 – RTUPB : PROFIL DE GUÉRISON 1/13

Le profil des patients, ci-dessus, montre un bénéfice rapide de l'opération, dès 3 mois, avec une qualité de miction (Qmax) qui progresse spectaculairement à 3 mois, et se stabilise à un régime de croisière à partir de 6 mois (17.0 ml/s).

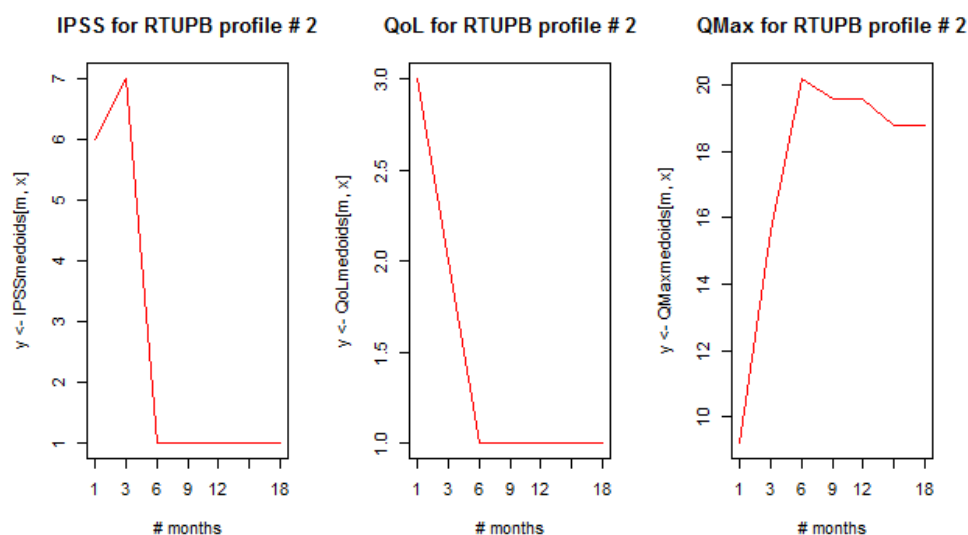


FIGURE 36 – RTUPB : PROFIL DE GUÉRISON 2/13

Le profil des patients, ci-dessus, montre un bénéfice progressif de l'opération, et stable à partir de 6 mois. La qualité de miction (Qmax) se stabilise à partir de 6 mois entre 18 et 20 ml/s.

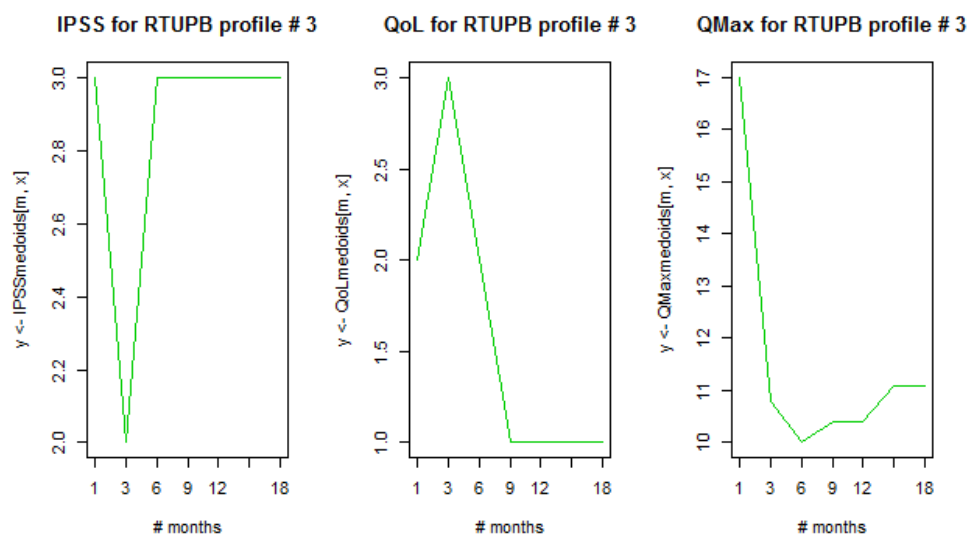


FIGURE 37 – RTUPB : PROFIL DE GUÉRISON 3/13

Le profil des patients, ci-dessus, montre un ressenti négatif à 3 mois (pic en hausse de QoL, pic en baisse de IPSS), et une dégradation significative de la qualité de miction (Qmax). L'indication QoL par contre redescend à partir de 9 mois. Peut-on interpréter ces résultats comme un résultat décevant de l'opération?

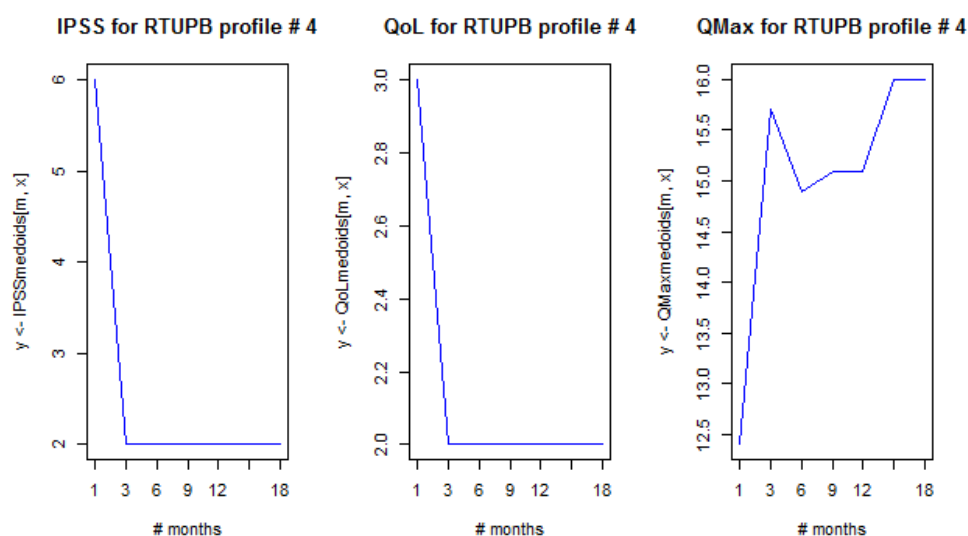


FIGURE 38 – RTUPB : PROFIL DE GUÉRISON 4/13

Le profil des patients, ci-dessus, est globalement assez proche de celui de la première classe (figure ??), avec une évolution différente de la qualité de miction (Qmax) avec des valeurs inférieures tout au long du suivi post-opératoire.

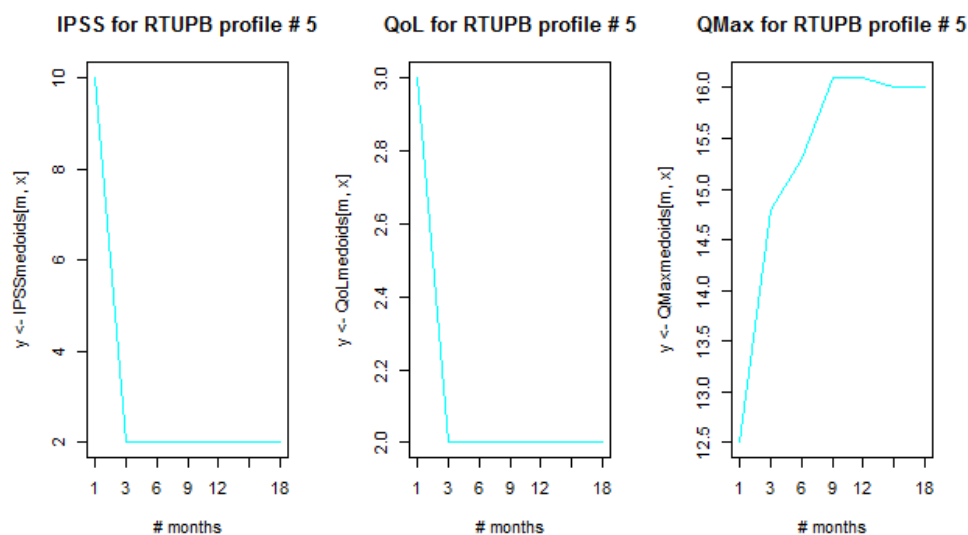


FIGURE 39 – RTUPB : PROFIL DE GUÉRISON 5/13

Ici aussi, le profil des patients ci-dessus, est assez proche de celui de la première classe (figure ??), avec une évolution différente de la qualité de miction (Qmax) plus progressive et sans pic à 3 mois.

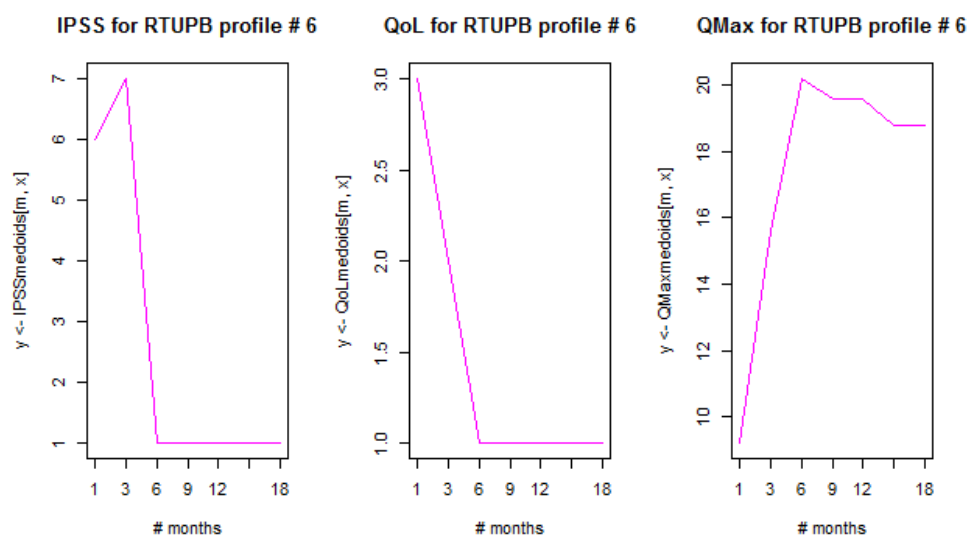


FIGURE 40 – RTUPB : PROFIL DE GUÉRISON 6/13

Le profil des patients, ci-dessus, est globalement très proche de celui de la deuxième classe (figure ??), avec des effets post-opératoires bénéfiques à partir de 6 mois.

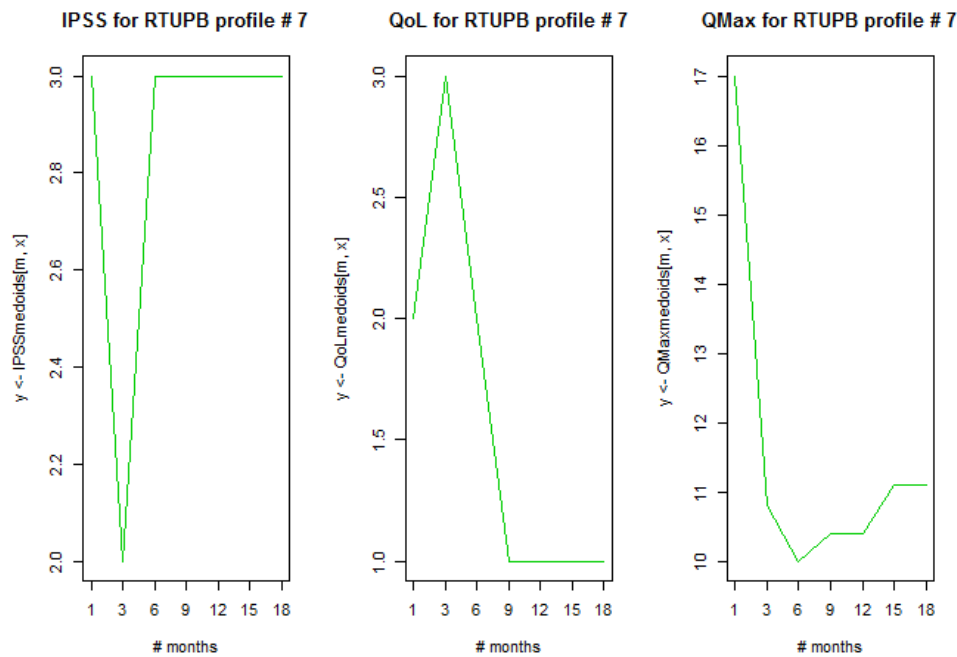


FIGURE 41 – RTUPB : PROFIL DE GUÉRISON 7/13

Le profil des patients, ci-dessus, est globalement très proche de celui de la troisième classe (figure ??), avec des résultats décevants.

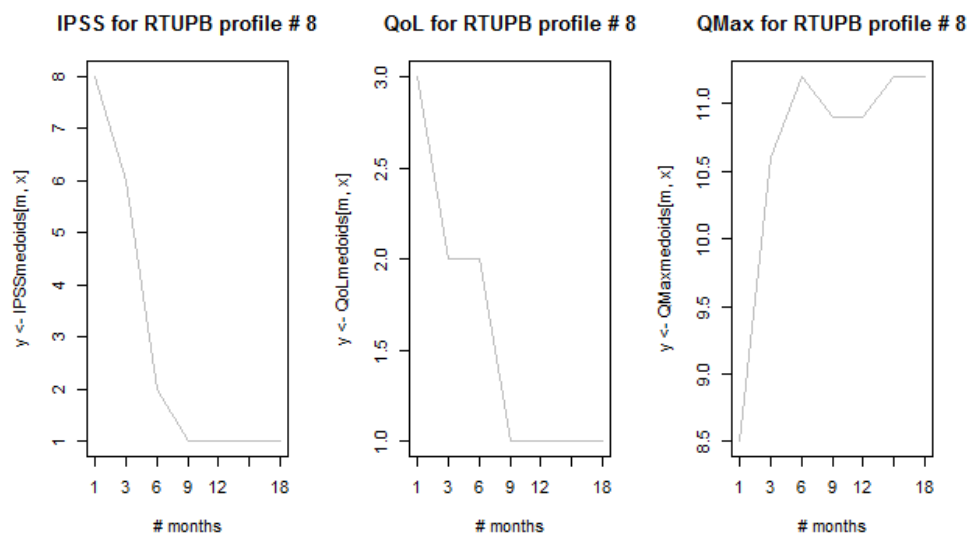


FIGURE 42 – RTUPB : PROFIL DE GUÉRISON 8/13

Le profil des patients, ci-dessus, montre un bénéfice de l'opération étalé sur les 9 premiers mois. La qualité de miction (Qmax) se stabilise à partir de 6 mois (11 ml/s).

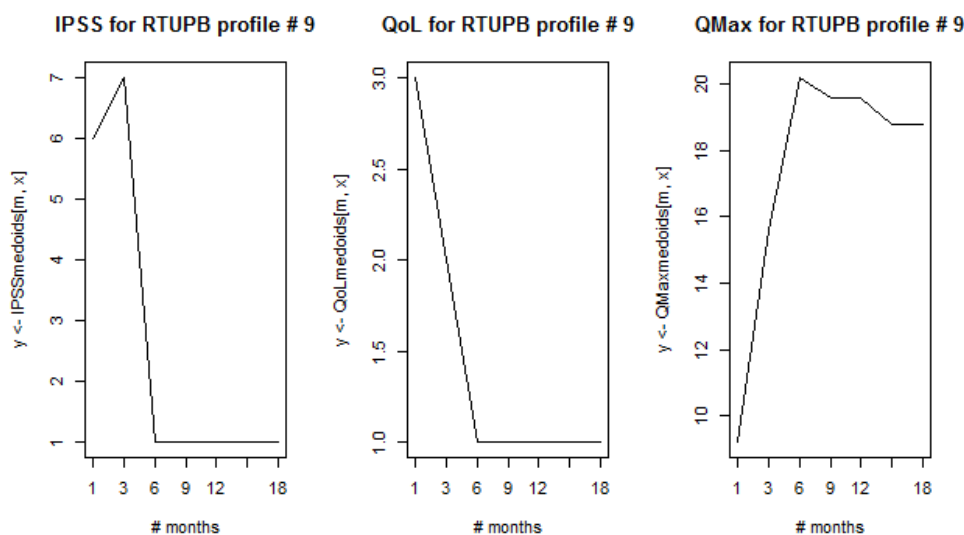


FIGURE 43 – RTUPB : PROFIL DE GUÉRISON 9/13

Le profil des patients, ci-dessus, est globalement très proche de celui de la deuxième classe (figure ??), avec des effets post-opératoires bénéfiques à partir de 6 mois.

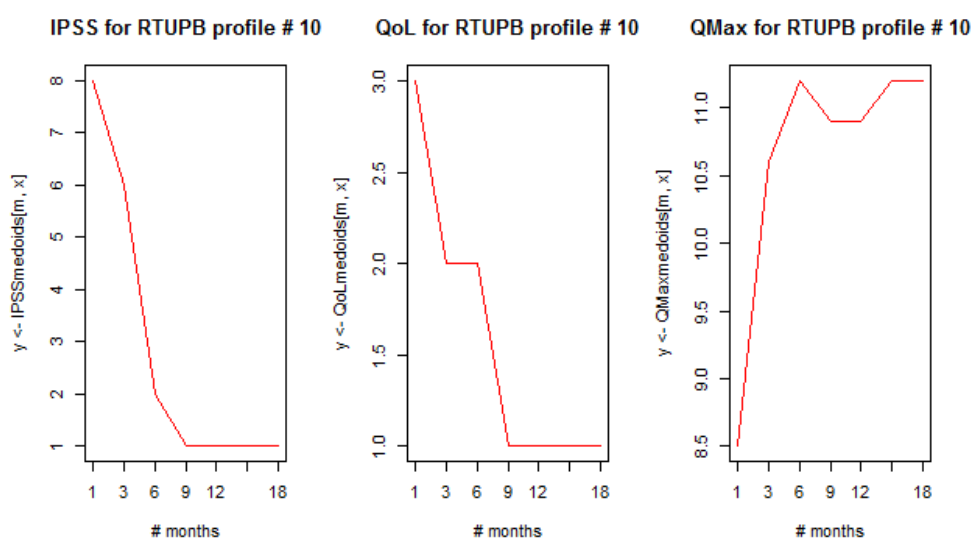


FIGURE 44 – RTUPB : PROFIL DE GUÉRISON 10/13

Le profil des patients, ci-dessus, très proche de celui de la huitième classe (figure ??), montre un bénéfice de l'opération étalé sur les 9 premiers mois. La qualité de miction (Qmax) se stabilise à partir de 6 mois (11 ml/s).

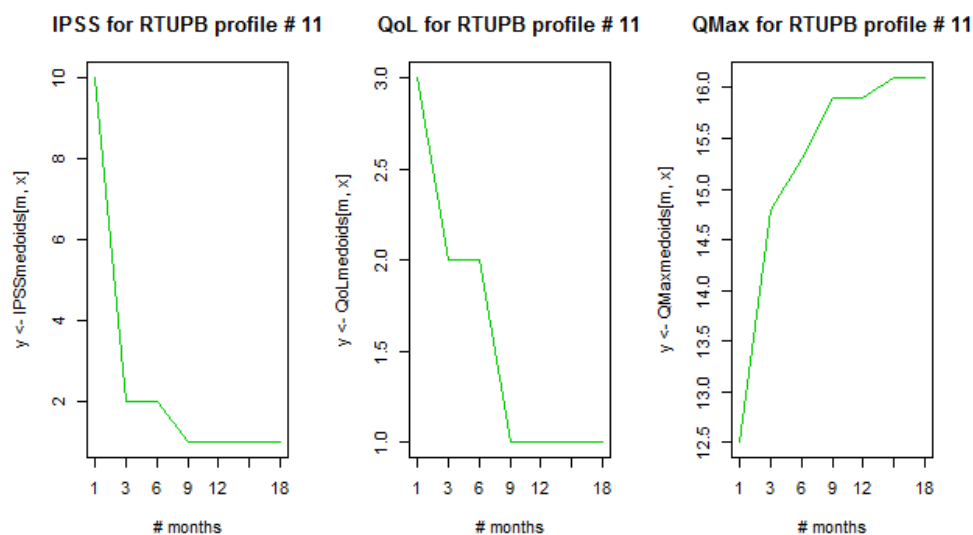


FIGURE 45 – RTUPB : PROFIL DE GUÉRISON 11/13

Le profil des patients, ci-dessus, assez proche de celui de la huitième classe (figure ??), montre un bénéfice de l'opération étalé sur les 9 premiers mois. La qualité de miction (Qmax) est au départ meilleure et se stabilise plus lentement après un an (11 ml/s).

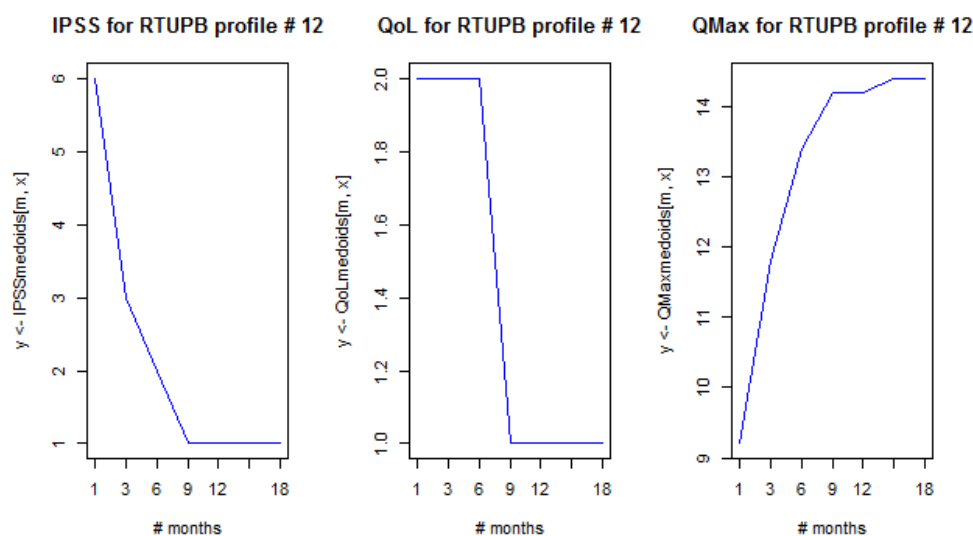


FIGURE 46 – RTUPB : PROFIL DE GUÉRISON 12/13

Le profil des patients, ci-dessus, assez proche de celui de la huitième classe (figure ??), montre un bénéfice de l'opération étalé sur les 9 premiers mois, avec une QoL qui ne descend qu'après 6 mois et une qualité de miction (Qmax) qui se stabilise plus lentement après un an (14 ml/s).

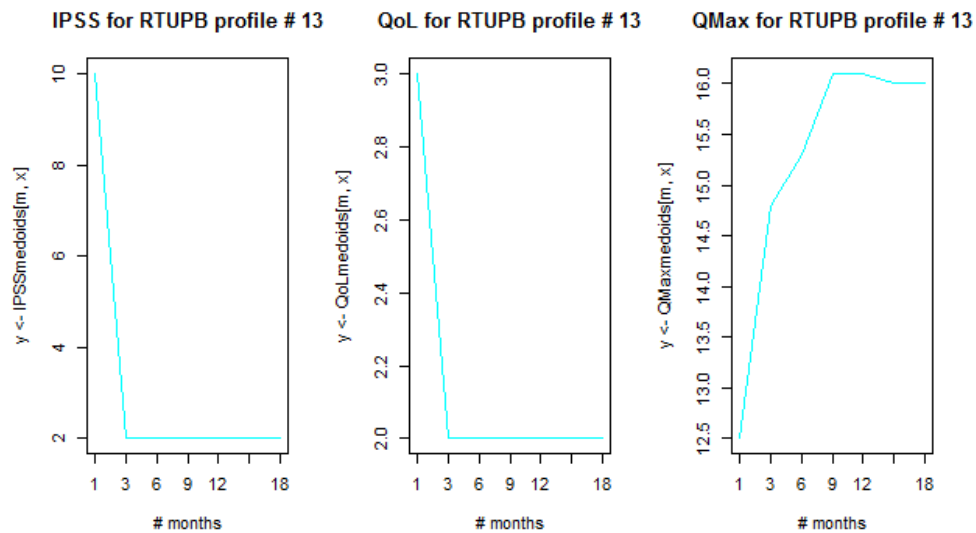


FIGURE 47 – RTUPB : PROFIL DE GUÉRISON 13/13

Le profil des patients, ci-dessus, est globalement assez proche de celui de la première classe (figure ??), avec une évolution différente de la qualité de miction (Qmax) avec des valeurs inférieures tout au long du suivi post-opératoire.

En résumé, il semble donc que d'un point de vue dynamique des profils de guérison, les 13 classes initiales se factorisent en 4 classes :

- résultat post-opératoire positif à partir de 3 mois (figure ??), regroupant 4 des 13 classes initiales,
- résultat post-opératoire positif à partir de 6 mois (figure ??), regroupant 3 des 13 classes initiales,
- résultat post-opératoire positif à partir de 9 mois (figure ??) regroupant 4 des 13 classes initiales,
- résultat post-opératoire décevant à partir de 3 mois (figure ??), regroupant 2 des 13 classes initiales,

4.7.2 VPPBS

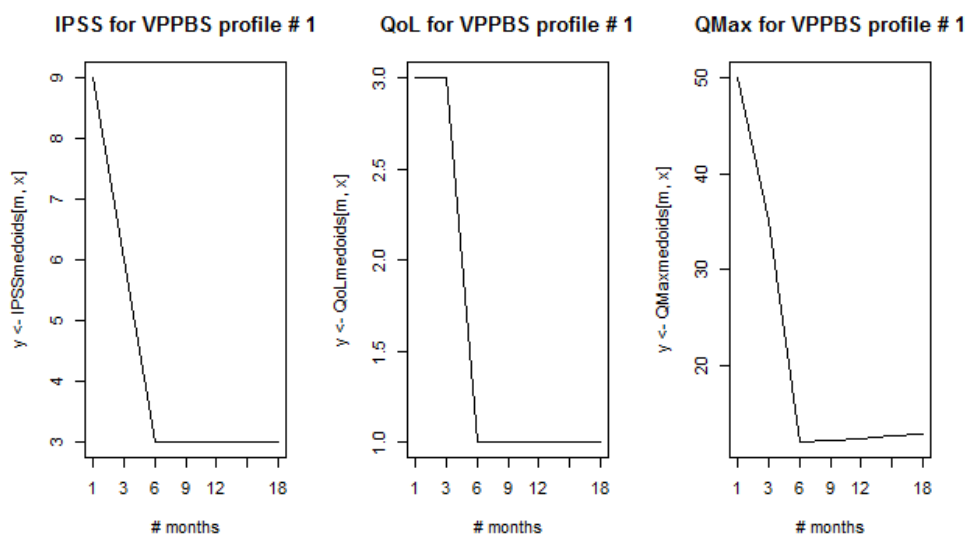


FIGURE 48 – VPPBS : PROFIL DE GUÉRISON 1/12

Le profil des patients, ci-dessus, montre une amélioration à 6 mois des indicateurs IPSS et QoL, mais en parallèle une dégradation de la qualité de la miction (Qmax). Tous les indicateurs restent ensuite stables au delà des 6 premiers mois.

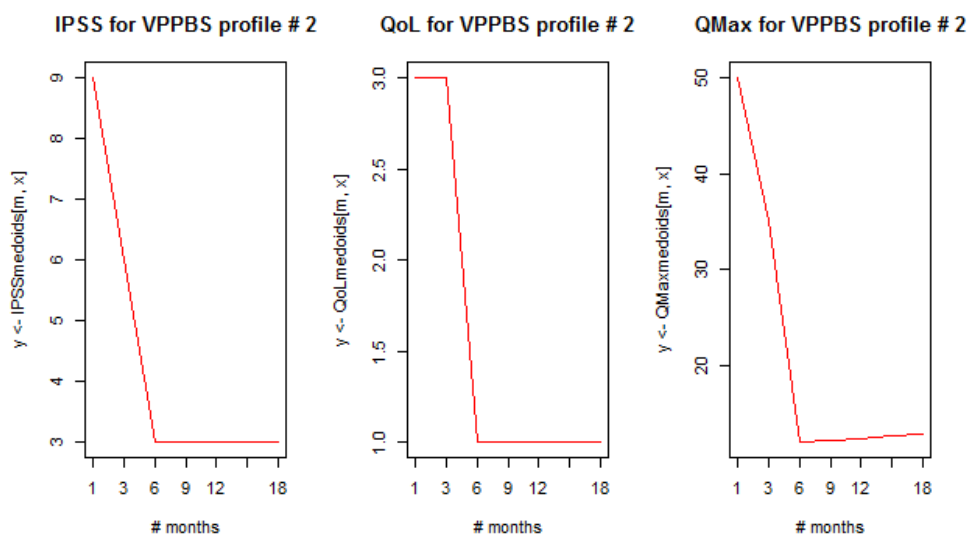


FIGURE 49 – VPPBS : PROFIL DE GUÉRISON 2/12

Le profil des patients, ci-dessus, est extrêmement très proche de celui de la première classe (figure ??).

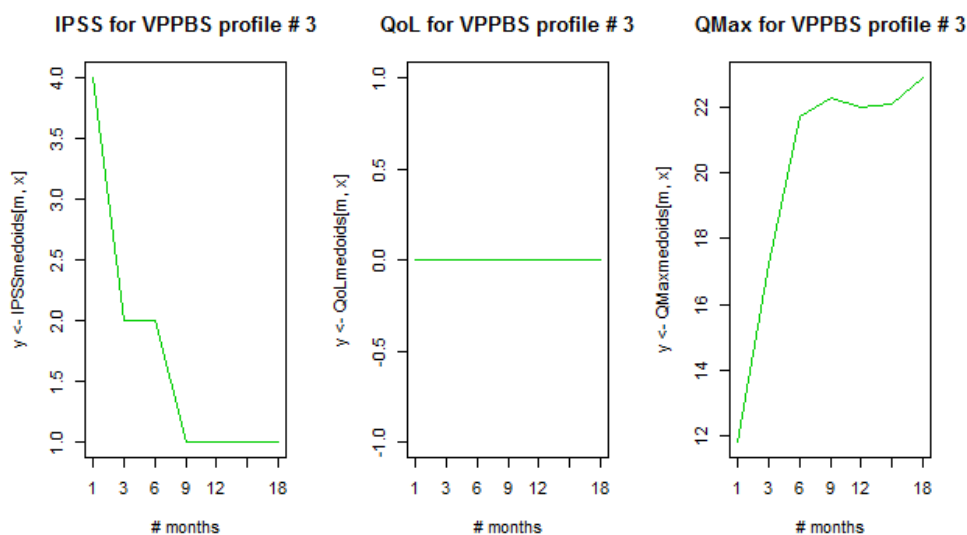


FIGURE 50 – VPPBS : PROFIL DE GUÉRISON 3/12

Le profil des patients, ci-dessus, montre une amélioration à 9 mois de l'indicateur IPSS (avec un plateau entre 3 et 6 mois), une stagnation de QoL (à 0 initialement, donc pas de souci), et une amélioration de la qualité de la miction (Qmax) très nette sur les 9 premiers mois, plus progressive voire quasi stable ensuite.

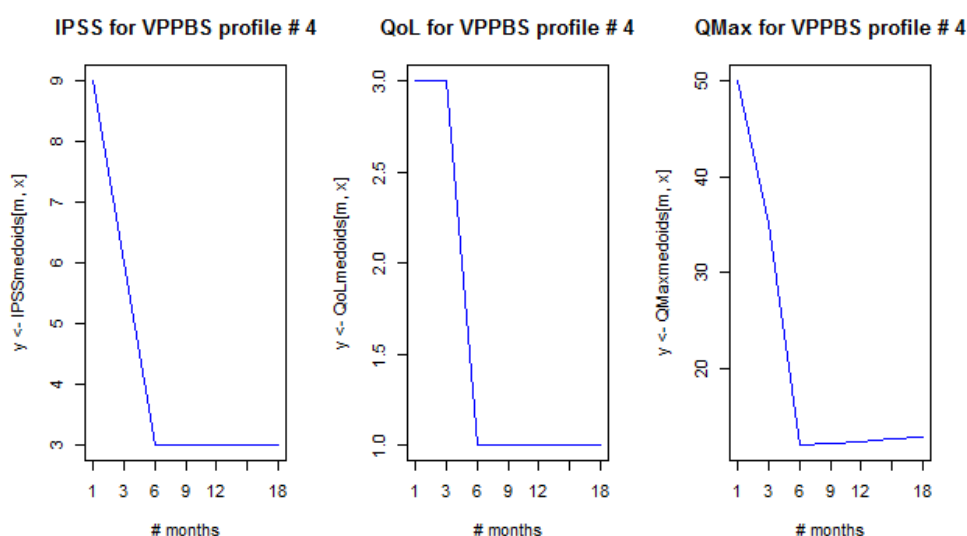


FIGURE 51 – VPPBS : PROFIL DE GUÉRISON 4/12

Le profil des patients, ci-dessus, est extrêmement très proche de celui de la première classe (figure ??).

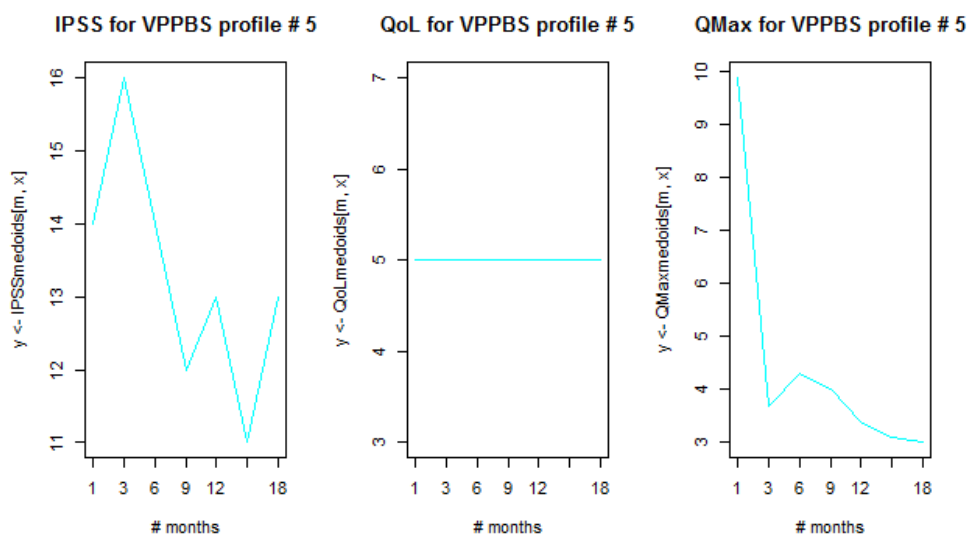


FIGURE 52 – VPPBS : PROFIL DE GUÉRISON 5/12

Le profil des patients, ci-dessus, montre une amélioration en dents de scie de l'indicateur IPSS (avec 1 pic à 3 mois), une stagnation de QoL à une valeur non nulle, et une dégradation de la qualité de la miction (Qmax) très nette sur les 3 premiers mois, puis plus progressive voire quasi stable ensuite.

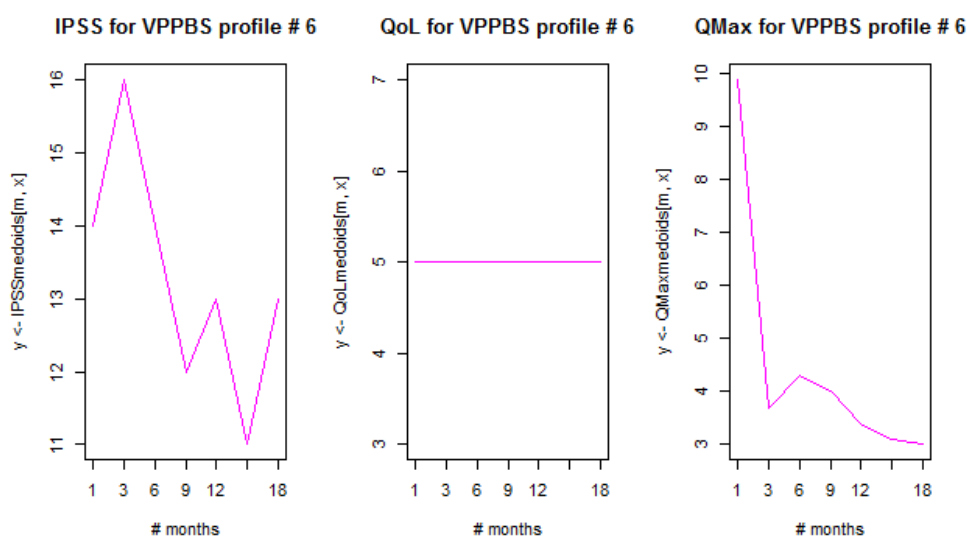


FIGURE 53 – VPPBS : PROFIL DE GUÉRISON 6/12

Le profil des patients, ci-dessus, est extrêmement très proche du précédent (figure ??).

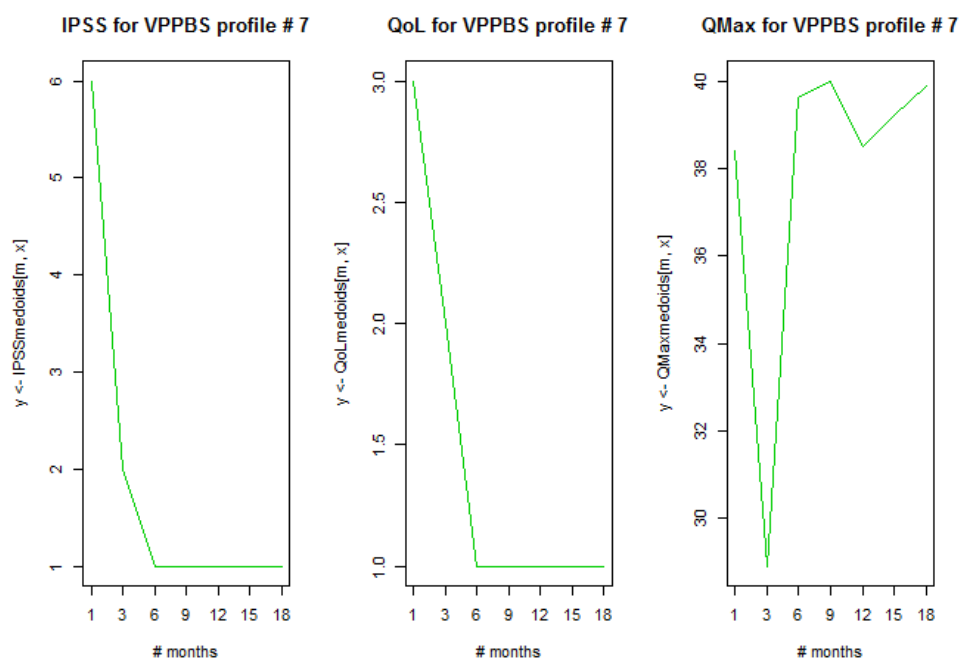


FIGURE 54 – VPPBS : PROFIL DE GUÉRISON 7/12

Le profil des patients, ci-dessus, montre une amélioration nette des indicateurs IPSS et QoL sur les 6 premiers mois. Ces indicateurs sont ensuite stables. La qualité de la miction (Qmax) s'améliore progressivement à 6 mois et au delà, après une forte chute à 3 mois.

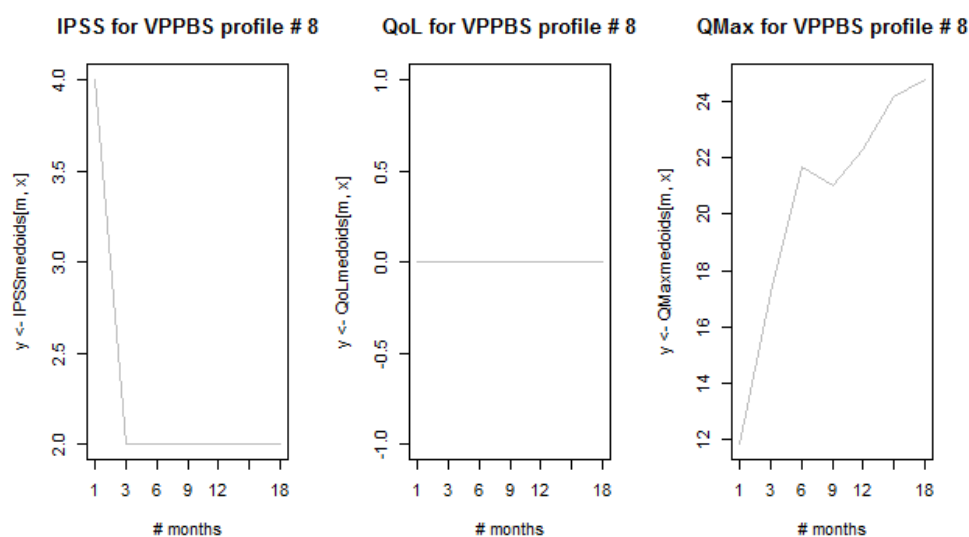


FIGURE 55 – VPPBS : PROFIL DE GUÉRISON 8/12

Le profil des patients, ci-dessus, est extrêmement assez proche de celui de la troisième classe (figure ??, mais nous noterons que la qualité de la miction (Qmax) continue ici à s'améliorer après les 6 premiers mois, jusqu'à la fin du suivi.

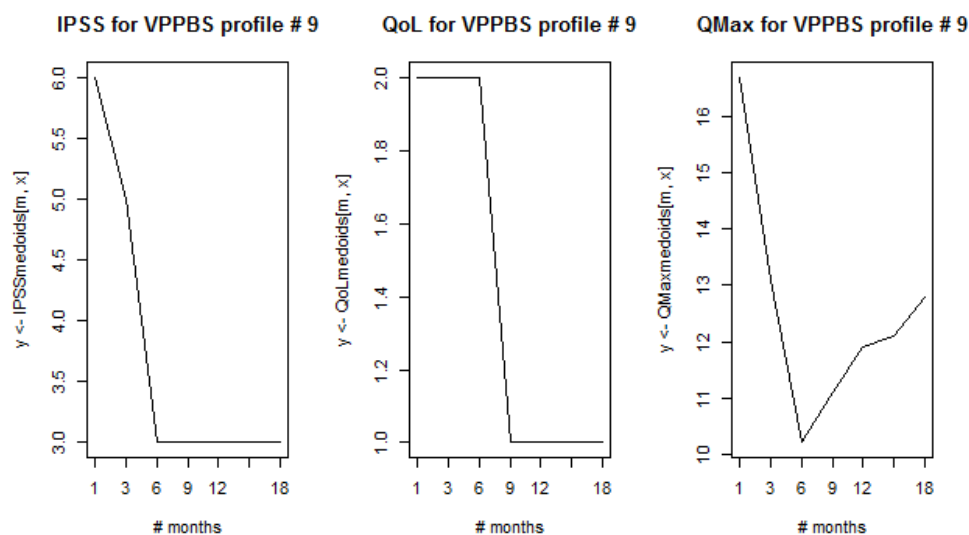


FIGURE 56 – VPPBS : PROFIL DE GUÉRISON 9/12

Le profil des patients, ci-dessus, est extrêmement très proche de celui de la première classe (figure ??).

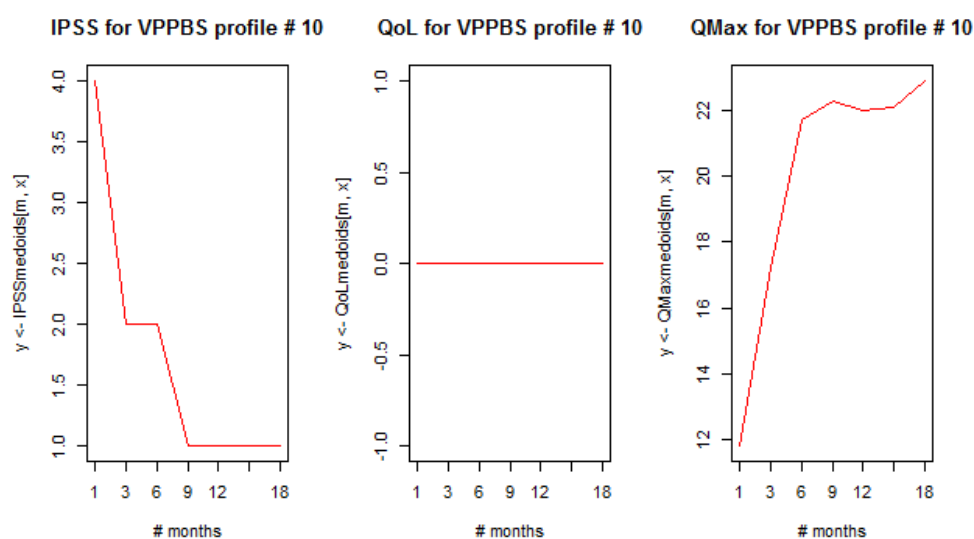


FIGURE 57 – VPPBS : PROFIL DE GUÉRISON 10/12

Le profil des patients, ci-dessus, est extrêmement assez proche de celui de la troisième classe (figure ??).

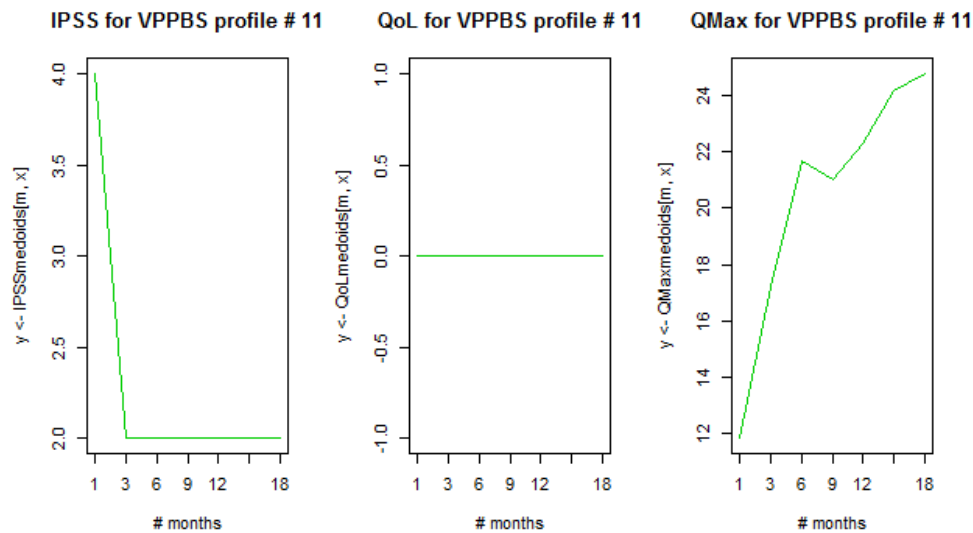


FIGURE 58 – VPPBS : PROFIL DE GUÉRISON 11/12

Le profil des patients, ci-dessus, est extrêmement très proche de celui de la huitième classe (figure ??).

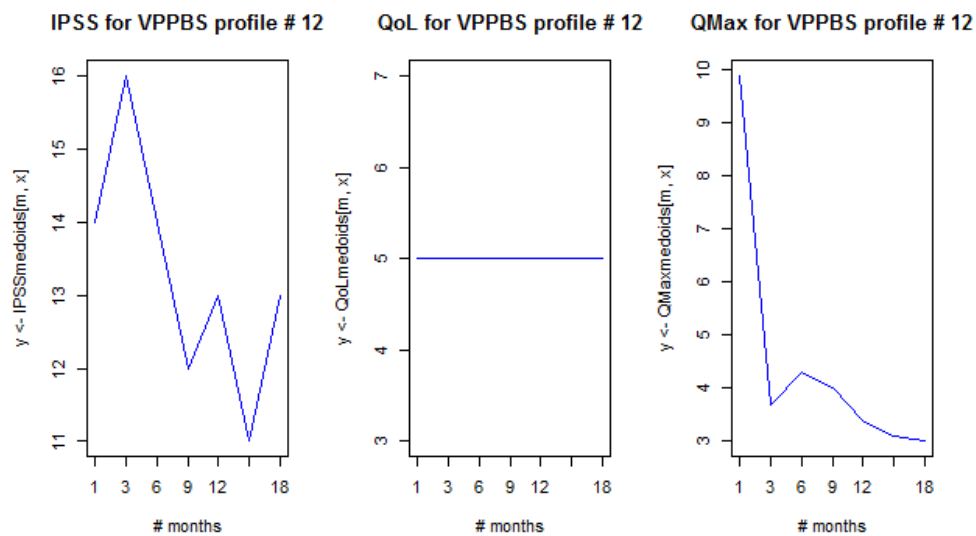


FIGURE 59 – VPPBS : PROFIL DE GUÉRISON 12/12

Le profil des patients, ci-dessus, est extrêmement très proche de celui de la cinquième classe (figure ??).

En résumé, il semble donc que d'un point de vue dynamique des profils de guérison, les 12 classes initiales se factorisent en 5 classes :

- résultat post-opératoire mitigé : meilleur IPSS/QoL, mais Qmax dégradé (figure ??), regroupant 4 des 12 classes initiales,
- résultat post-opératoire positif dès 3 mois, quasi stable après 9 mois : meilleur IPSS, QoL inchangé, meilleur Qmax (figure ??), regroupant 2 des 12 classes initiales,

- résultat post-opératoire positif dès 3 mois, qui continue à s'améliorer après 9 mois : meilleur IPSS, QoL inchangé, meilleur Qmax (figure ??), regroupant 2 des 12 classes initiales,
- résultat post-opératoire mitigé : amélioration faible et en dents de scie d'IPSS, QoL inchangé, Qmax dégradé à partir de 3 mois (figure ??), regroupant 3 des 12 classes initiales,
- résultat post-opératoire positif à 6 mois, avec une légère amélioration de Qmax après une forte chute à 3 mois (figure ??), regroupant 2 des 12 classes initiales,

4.8 LIENS ENTRE PROFILS PRÉ ET POST-OPÉRATOIRES

4.8.1 LIENS ENTRE PROFILS PRÉ ET POST-OPÉRATOIRES

Pour comparer les résultats précédents et les profils identifiés, nous reprenons et rapprochons pour chaque patient son profil pré-opératoire et son profil post-opératoire. La liste des couples de profils ainsi obtenue est triée

- par profil pré-opératoire : figure ?? et figure ?? pour respectivement RTUPB et VPPBS
- par profil post-opératoire : figure ?? et figure ?? pour respectivement RTUPB et VPPBS

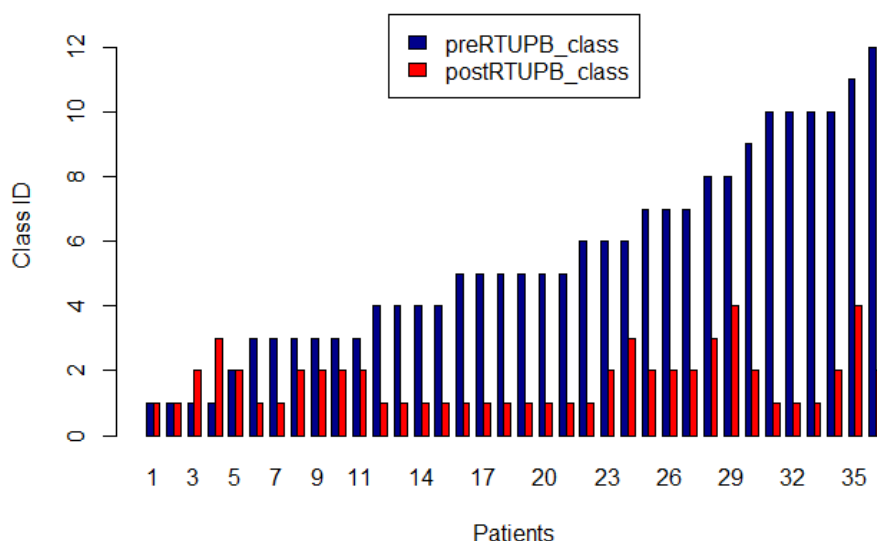


FIGURE 60 – RTUPB : CLASSES POST-OPÉRATOIRES EN FONCTION DES CLASSES PRÉ-OPÉRATOIRES

La figure ?? révèle que pour les patients appartenant au profil pré-opératoire 4, 5 et 7 le résultat constaté est toujours le profil post-opératoire n°1. Pour les autres profils pré-opératoires (à l'exception des singletons pour lesquels les résultats ne sont donc pas confirmés), il n'existe pas de profils post-opératoires uniques permettant d'en déduire une prévision de guérison. La figure ?? n'apporte pas d'informations complémentaires.

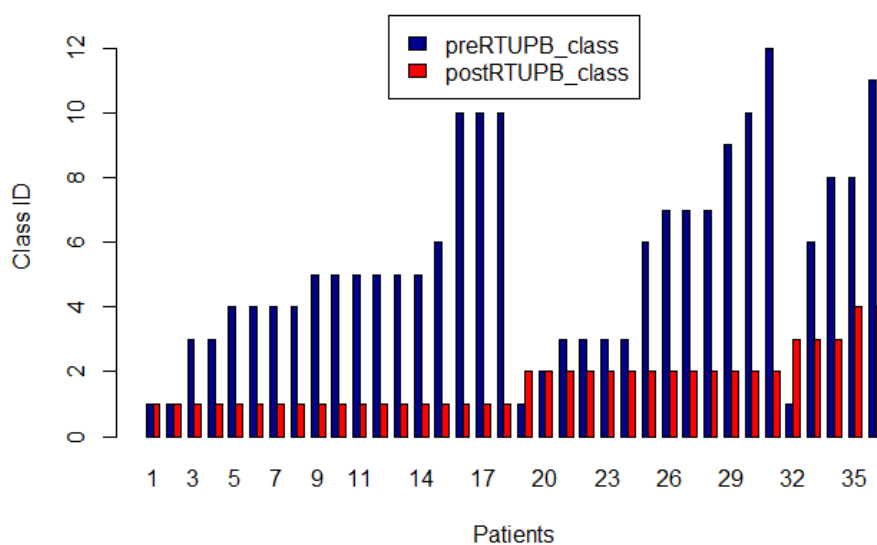


FIGURE 61 – RTUPB : CLASSES PRÉ-OPÉRATOIRES EN FONCTION DES CLASSES POST-OPÉRATOIRES

Pour VPPBS, la figure ?? montre la correspondance d'un seul profil post-opératoire pour chaque profil pré-opératoire. Nous pouvons donc nous baser sur le profil pré-opératoire pour envisager ou non une opération VPPBS et prévoir la guérison.

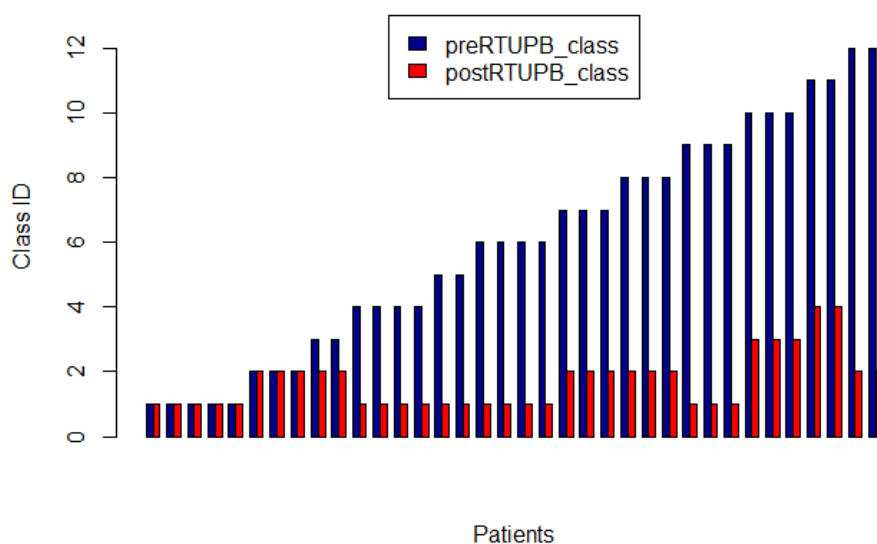


FIGURE 62 – VPPBS : CLASSES POST-OPÉRATOIRES EN FONCTION DES CLASSES PRÉ-OPÉRATOIRES

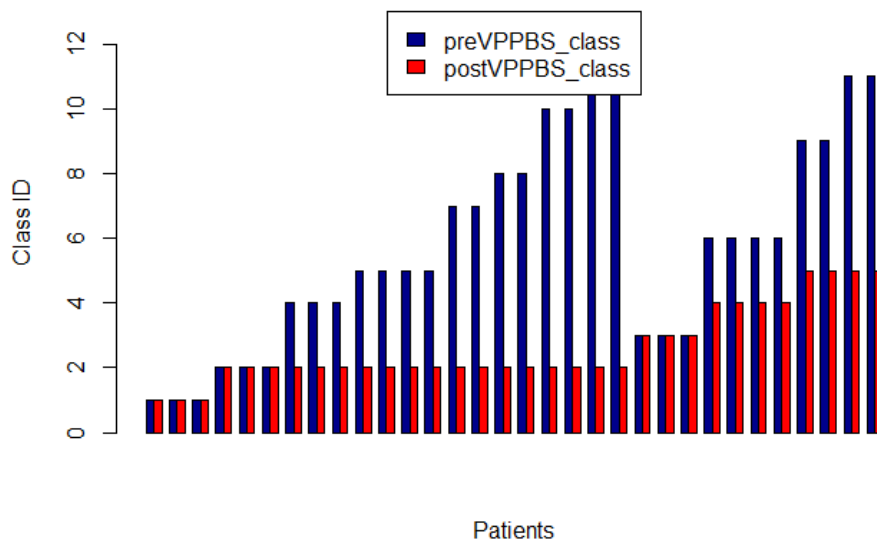


FIGURE 63 – VPPBS : CLASSES PRÉ-OPÉRATOIRES EN FONCTION DES CLASSES POST-OPÉRATOIRES

5 CLASSIFICATION SUPERVISÉE RTUPB-VPPBS ET VAPOR

5.1 RTUPB : PRÉDICTION IPSS, QOL ET QMAX À 12 MOIS

Dans les sections qui suivent, nous réutilisons la même méthode pour construire un arbre de régression : pour chacune des techniques opératoires (respectivement RTUPB, VPPBS et VAPOR) et à partir des données pré-opératoires précédemment utilisées et nettoyées des variables invariantes ou incomplètes, auxquelles on ajoutera la variable à prédire, nous effectuons un tirage aléatoire pour créer deux ensembles de données :

- un ensemble d'apprentissage représentant 80% des patients opérés par une technique. A partir de cet ensemble, nous inférerons un arbre de régression en utilisant le package `rpart` sous R.
- un ensemble de validation représentant 20% des patients opérés par la même technique.

Les variables IPSS et QoL seront traitées comme des variables ordinales, tandis que la variable Qmax sera traitée comme une variable linéaire.

Pour chaque variable et chaque base, nous comparerons les résultats obtenus par un arbre de régression (inféré par élagage d'un arbre "complet") et les résultats obtenus par une forêt aléatoire d'arbres de régression.

5.1.1 RTUPB : IPSS À 12 MOIS

L'arbre de régression complet obtenu pour la variable IPSS à 12 mois (colonne IPSS_4) met en œuvre 6 variables (Cf. figure ??). Les feuilles de l'arbre représentent, en première ligne, la valeur prédite (la plus probable) pour IPSS à 12 mois. La ligne suivante indique les probabilités pour chacun des 3 niveaux constatés sur l'échantillon d'apprentissage pour IPSS à 12 mois, c'est-à-dire respectivement 1/2/3. Par exemple en cas de caillotage (valeur == 1), l'arbre de régression conduit à la feuille en bas à gauche et prédit une valeur IPSS à 1 avec une probabilité de 88%.

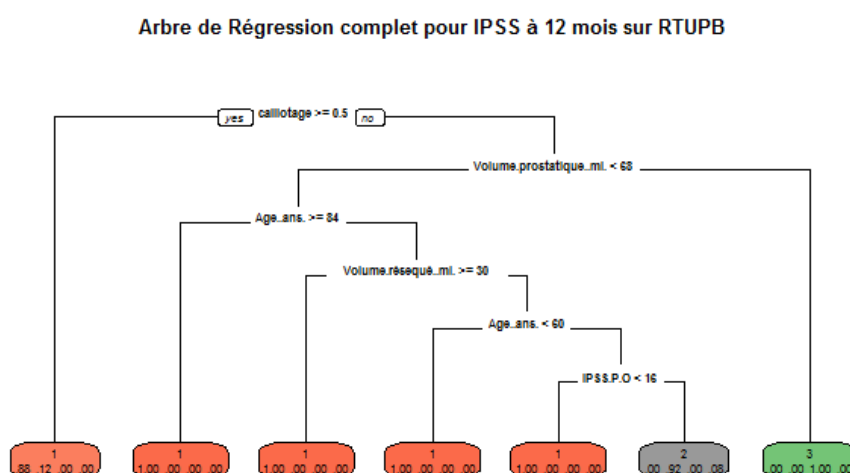


FIGURE 64 – RTUPB : ARBRE DE RÉGRESSION POUR IPSS À 12 MOIS

Afin de simplifier l'arbre, nous vérifions le facteur d'amélioration (complexity parameter) apporté par chaque embranchement de l'arbre, figure ??). Nous élaguons l'arbre au niveau du coude d'inflexion de la courbe, soit environ à $cp=0.13$, pour obtenir un arbre simplifié (figure ??).

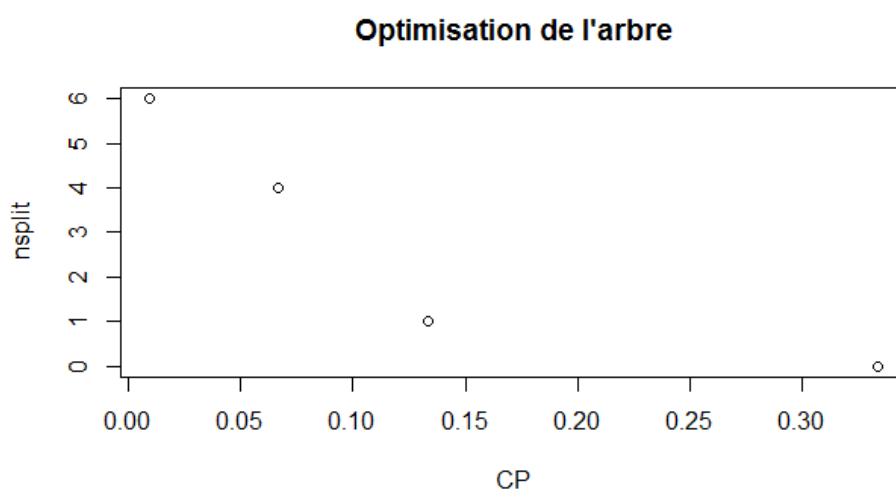


FIGURE 65 – RTUPB : OPTIMISATION DE L'ARBRE DE RÉGRESSION POUR IPSS À 12 MOIS

En appliquant cet arbre élagué (figure ??) sur l'ensemble de tests, nous obtenons pour chaque patient une table de probabilités pour chacune des valeurs IPSS 1/2/3/4. Les prédictions (valeurs les plus probables) sont correctes pour tous les patients à l'exception des patients 20 et 29 pour lesquels la valeur référence IPSS à 12 mois est de 1. Nous obtenons la table de correspondance suivante, entre valeurs de référence et valeurs de prédiction :

	prediction			
validation	1	2	3	4
	1	0	2	0
	2	0	5	1
	3	0	0	0
	4	0	0	0

Soit un taux d'erreur de 37,5%.

Arbre de Régression élagué pour IPSS à 12 mois sur RTUPB

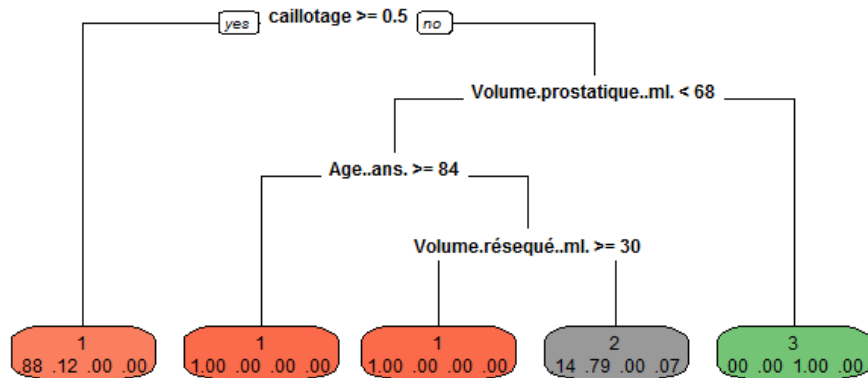


FIGURE 66 – RTUPB : ARBRE DE RÉGRESSION POUR IPSS À 12 MOIS

Prédiction IPSS pour échantillon de test RTUPB

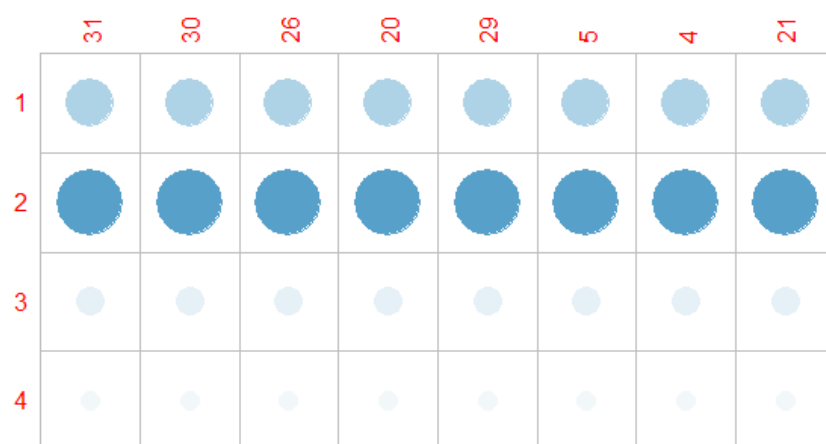


FIGURE 67 – RTUPB : PRÉVISION POUR IPSS À 12 MOIS

En utilisant une forêt aléatoire d'arbres de régression, nous obtenons une nouvelle table de correspondance :

	prediction				
validation	1	2	3	4	
	1	1	1	0	0
	2	0	4	1	1

Soit un taux d'erreur de 37,5%. Sous réserve qu'une marge d'erreur de prédiction d'un point est acceptable sur l'indice IPSS (hypothèse à valider par un avis médical), le taux d'erreur serait alors ramené à 12,5%.

5.1.2 RTUPB : QOL À 12 MOIS

L'arbre de régression obtenu pour la variable QoL à 12 mois (colonne QoL__4) met en œuvre cinq variables (Cf. figure ??). Les feuilles de l'arbre représentent, en première ligne, la valeur prédite (la plus probable) pour QoL à 12 mois. La ligne suivante indique les probabilités pour chacun des 3 niveaux constatés sur l'échantillon d'apprentissage pour QoL à 12 mois, c'est-à-dire respectivement 0/1/2.

Arbre de Régression élagué pour QoL à 12 mois sur RTUPB

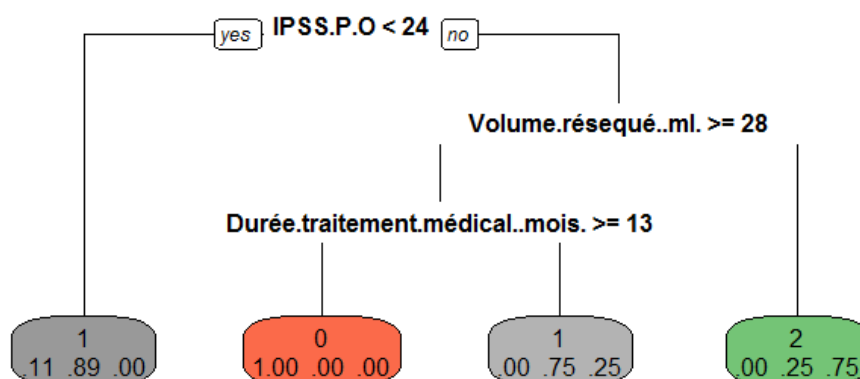


FIGURE 68 – RTUPB : ARBRE DE RÉGRESSION COMPLET POUR QOL À 12 MOIS

Nous vérifions l'amélioration apportée par chaque embranchement de cet arbre de régression complet, figure ?? . Un élagage au point d'inflexion ($cp=0.15$) permet d'obtenir un arbre simplifié, représenté figure ??.

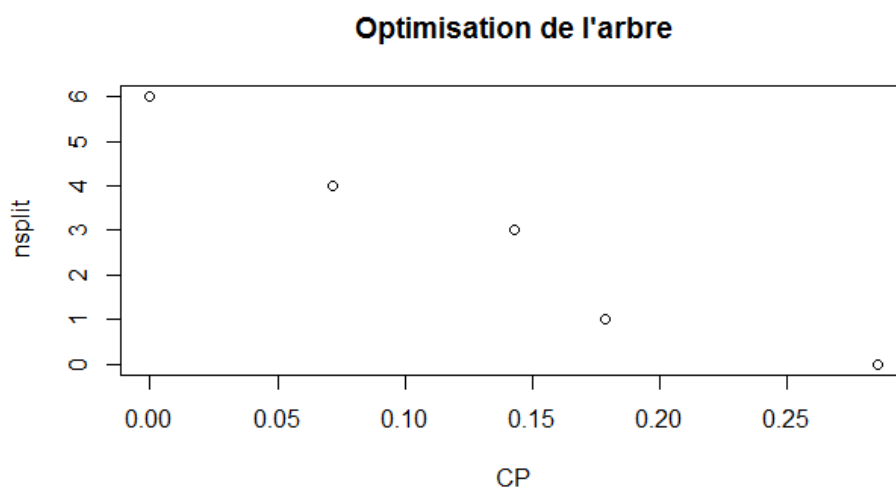


FIGURE 69 – RTUPB : OPTIMISATION DE L'ARBRE DE RÉGRESSION POUR QOL À 12 MOIS

Arbre de Régression élagué pour QoL à 12 mois sur RTUPB

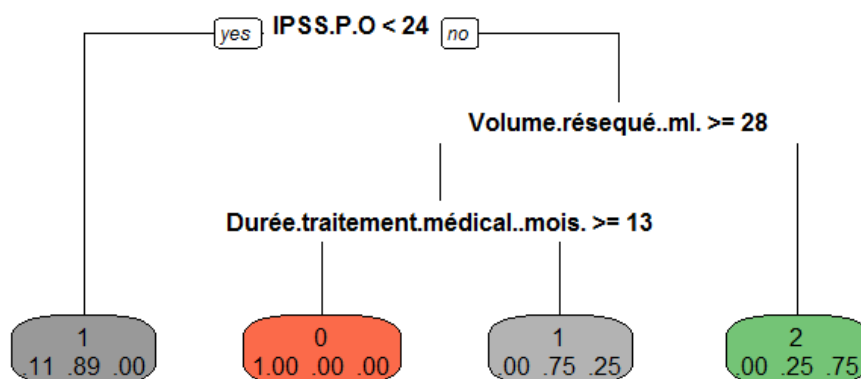


FIGURE 70 – RTUPB : ARBRE DE RÉGRESSION ÉLAGUÉ POUR QOL À 12 MOIS

Prédiction QoL pour échantillon de test RTUPB

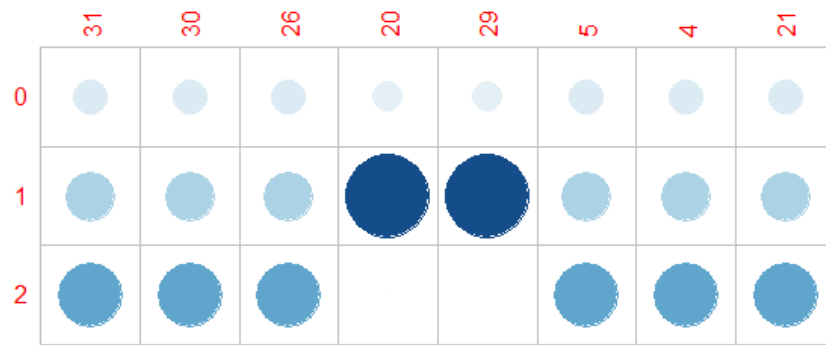


FIGURE 71 – RTUPB : PRÉVISION POUR QOL À 12 MOIS

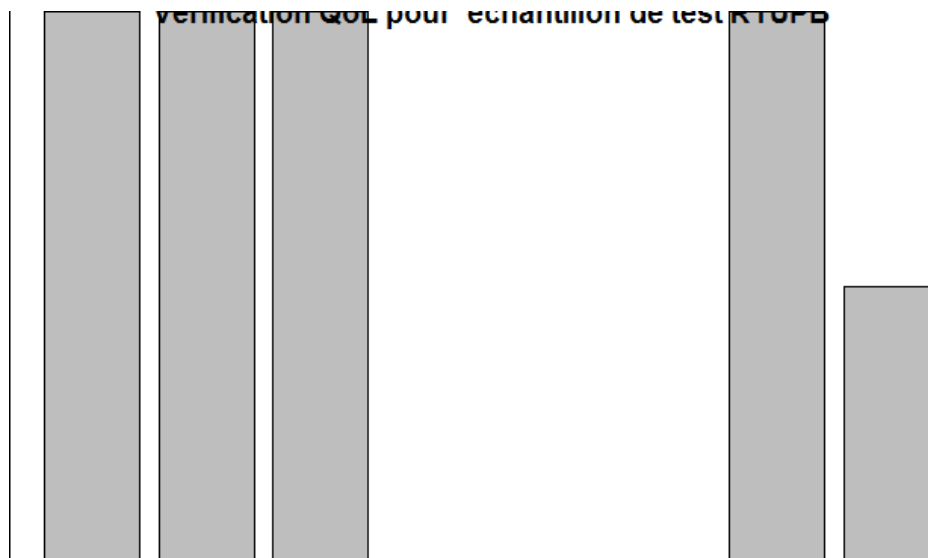


FIGURE 72 – RTUPB : VALEURS TEST POUR QOL À 12 MOIS

En appliquant cet arbre sur l'ensemble de tests, nous obtenons pour chaque patient une table de probabilités (figure ??) pour chacune des valeurs QoL 0/1/2. On peut alors comparer les prédictions (les valeurs les plus probables) avec les valeurs de référence (figure ??). Celles-ci sont correctes pour 4 patients sur 8, et incorrectes pour les 4 autres (20, 29, 5 et 21). Nous obtenons la table de correspondance suivante, entre valeurs de référence et valeurs de prédiction :

	prediction		
validation	0	1	2
0	0	2	1
1	0	1	0
2	0	0	4

Soit un taux d'erreur de 37,5%.

Afin d'améliorer les résultats de prédiction, il nous faudrait répéter l'opération et utiliser une forêt d'arbres de régression. La prédiction utilisée sera alors la prédiction moyenne de l'ensemble des arbres de la forêt. Nous obtenons alors une nouvelle table de correspondance :

	prediction			
validation	0	1	2	
0	1	1	1	
1	0	1	0	
2	0	0	4	

Soit un taux d'erreur moindre : 25%.

5.1.3 RTUPB : QMAX À 12 MOIS

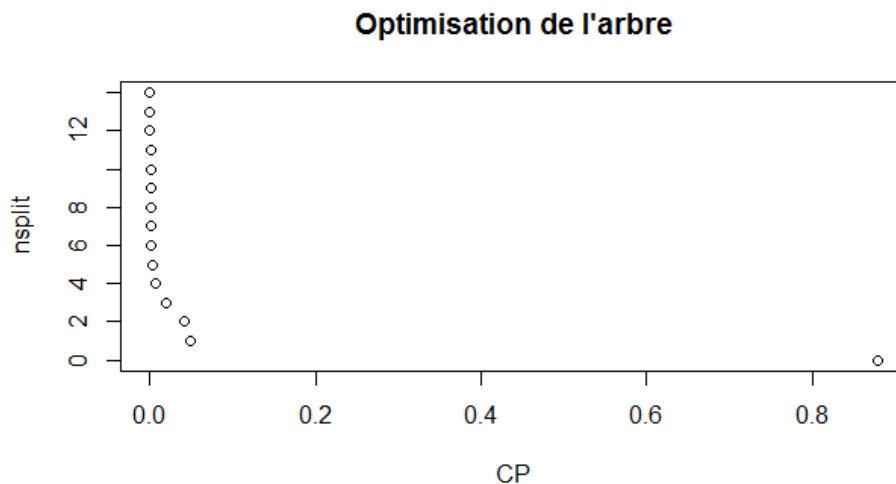


FIGURE 73 – RTUPB : OPTIMISATION DE L'ÉLAGAGE DE L'ARBRE COMPLET POUR QMAX À 12 MOIS

L'arbre de régression obtenu pour la variable Qmax à 12 mois (colonne Qmax (ml/s)_3), après élagage (Cf. figure pour le choix d'élagage optimal), met en œuvre 2 variables (Cf. figure ??). Les feuilles de l'arbre représentent, en première ligne, la valeur approximative pour Qmax à 12 mois. La ligne suivante indique le nombre de patients de l'échantillon d'apprentissage correspondant à cette feuille.

Arbre de Régression élagué pour Qmax à 12 mois sur RTUPB

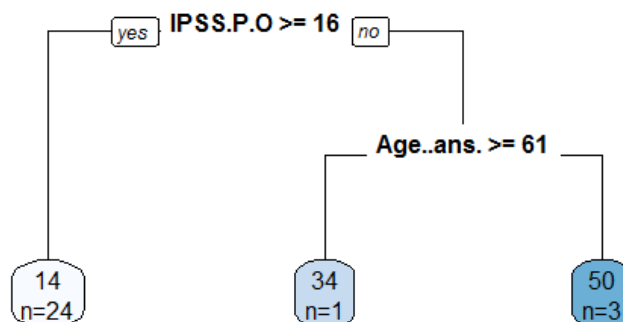


FIGURE 74 – RTUPB : ARBRE DE RÉGRESSION POUR QMAX À 12 MOIS

Pour évaluer l'arbre de régression, nous utiliserons une mesure de distance (valeur absolue de la différence) pour entre la valeur prédite et la valeur de référence pour chaque patient de l'ensemble de validation. La figure ?? représente l'écart constaté entre valeurs prédites (colonne de gauche) et valeurs de référence (colonne de droite). La moyenne de ces écarts nous donne une mesure du taux d'erreur de l'arbre. En répétant l'opération pour construire une forêt d'arbres de régression, nous utiliserons alors la prédiction rendue par l'arbre avec le plus faible taux d'erreur (Cf. figure ??).

Vérification Qmax pour échantillon de test RTUPB

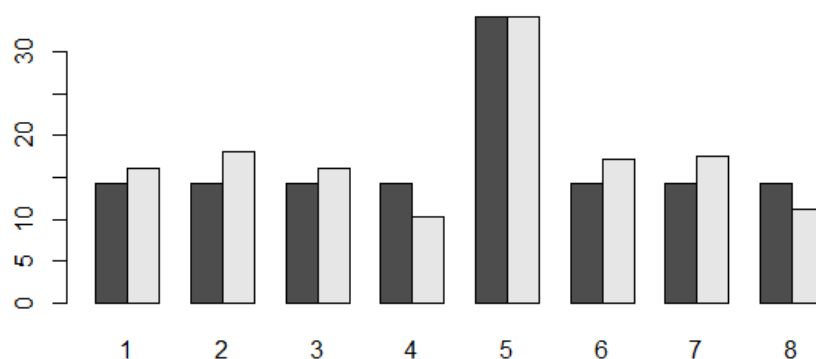


FIGURE 75 – RTUPB : EVALUATION DES PRÉVISIONS POUR QMAX À 12 MOIS

Vérification Qmax (RandomForest) pour échantillon de test RTUPE

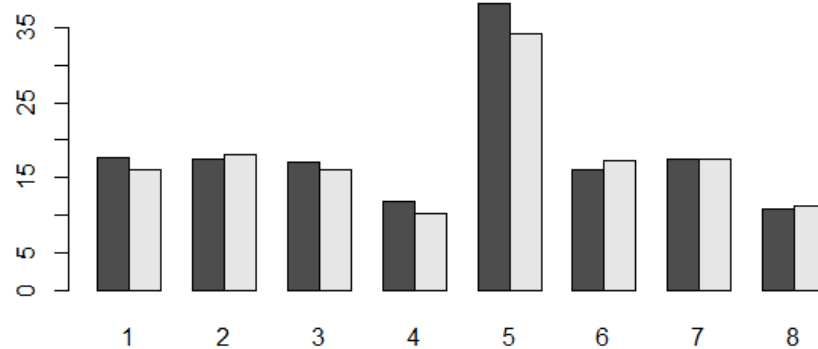


FIGURE 76 – RTUPB : EVALUATION DES PRÉVISIONS D'UNE FORÊT ALÉATOIRE POUR QMAX À 12 MOIS

5.2 VPPBS : PRÉDICTION IPSS, QOL ET QMAX À 12 MOIS

5.2.1 VPPBS : IPSS À 12 MOIS

Nous appliquons la même méthode que précédemment sur les données VPPBS. Nous obtenons un arbre de régression sur la seule variable PSA (figure ??) qui indique sur chaque feuille la probabilité des valeurs respectives d'IPSS à 12 mois : 0/1/2/3/4/5/13, constatées sur l'ensemble d'apprentissage.

Arbre de Régression pour IPSS à 12 mois sur VPPBS

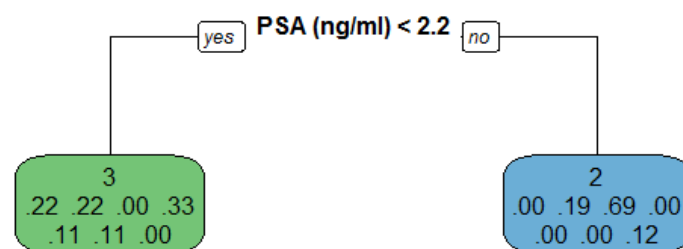


FIGURE 77 – VPPBS : ARBRE DE RÉGRESSION POUR IPSS À 12 MOIS

La figure ?? représente plus visuellement les probabilités prédites par cet arbre de régression pour l'ensemble de validation. On s'aperçoit que la plus forte probabilité qui ressort est

quasi systématiquement $IPSS=2$, alors que cette valeur correspond à la valeur de référence pour le patient 40 seulement. Cela est dû au tirage aléatoire qui a suréchantillonné dans l'ensemble d'apprentissage les patients pour lesquels la valeur IPSS à 12 mois vaut 2, et sous échantillonné les patients pour lesquels la valeur IPSS à 12 mois vaut 3,4,5 ou 13.

Nous obtenons la table de correspondance suivante, entre valeurs de référence et valeurs de prédiction :

	prediction						
validation	0	1	2	3	4	5	13
1	0	0	0	1	0	0	0
2	0	0	1	0	0	0	0
3	0	0	3	0	0	0	0
4	0	0	0	1	0	0	0
13	0	0	1	0	0	0	0

Soit un taux d'erreur de 85,7%.

Répéter la génération d'arbre de régression pour construire une forêt permettrait de minimiser ce risque d'erreur. Nous obtenons alors une nouvelle table de correspondance :

	prediction						
validation	0	1	2	3	4	5	13
1	0	1	0	0	0	0	0
2	0	1	0	0	0	0	0
3	0	0	3	0	0	0	0
4	0	0	0	1	0	0	0
13	0	0	0	0	0	0	1

Soit un taux d'erreur de 71,4%. Si toutefois, une marge d'erreur d'un point est médicalement raisonnable, alors le taux d'erreur est quasi nul.

Prédiction IPSS pour échantillon de test VPPBS

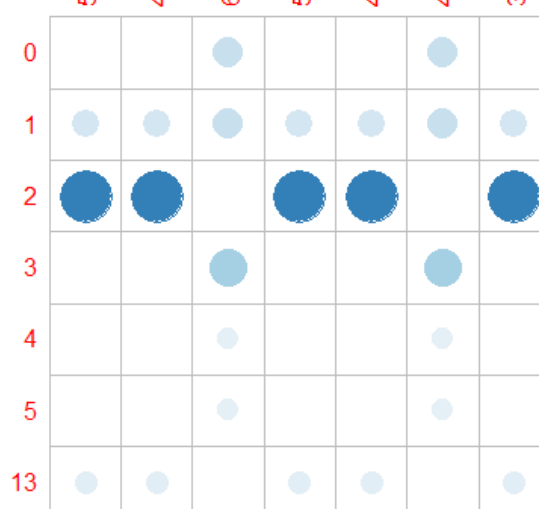


FIGURE 78 – VPPBS : PRÉVISION POUR IPSS À 12 MOIS

5.2.2 VPPBS : QOL À 12 MOIS

On retrouve la variable PSA dans cet arbre de régression pour la QoL à 12 mois qui infère des probabilités pour les valeurs respectives 0/1/3/5 (Cf. figure ??). Le tirage aléatoire a ici exclu la valeur 2 (patient 60 uniquement) de l'ensemble d'apprentissage. La validation pour cette valeur sera donc biaisée (Cf. figure ??).

La table de correspondance entre valeurs de référence et valeurs de prédiction est la suivante :

	prediction				
validation	0	1	3	5	
0	2	0	0	0	
1	2	1	0	0	
2	0	1	0	0	
5	1	0	0	0	

Soit un taux d'erreur de 57%.

Arbre de Régression pour QOL à 12 mois sur VPPBS

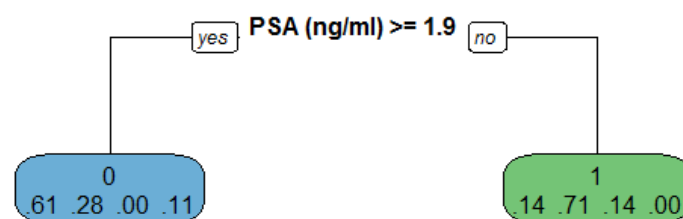


FIGURE 79 – VPPBS : ARBRE DE RÉGRESSION POUR QOL À 12 MOIS

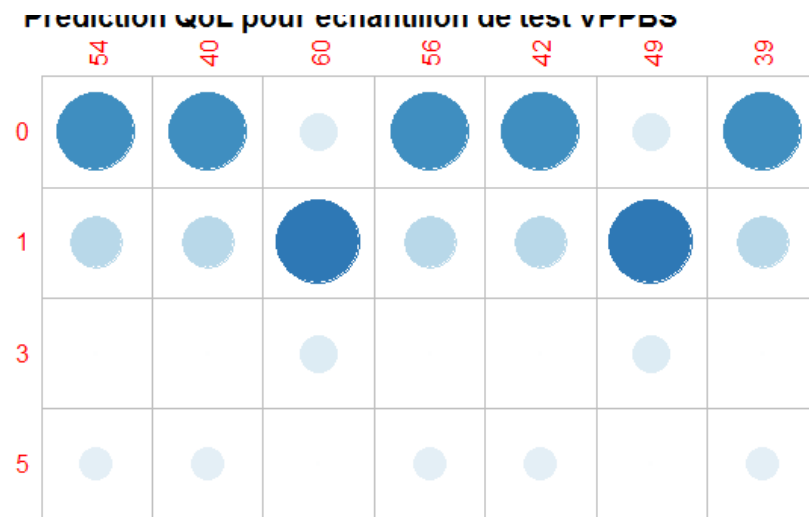


FIGURE 80 – VPPBS : PRÉVISION POUR QOL À 12 MOIS

Répéter la génération d'arbre de régression pour construire une forêt permet de minimiser ce risque d'erreur, et produit une nouvelle table de correspondance :

	prediction				
validation	0	1	3	5	
	0	2	0	0	0
	1	2	1	0	0
	2	0	1	0	0
	5	0	0	0	1

Soit un taux d'erreur de 43%.

5.2.3 VPPBS : QMAX À 12 MOIS

L'arbre de régression obtenu pour la variable Qmax à 12 mois (colonne Qmax (ml/s)_3) met en œuvre une seule variable (Cf. figure ??). Les feuilles de l'arbre représentent, en première ligne, la valeur approximative pour Qmax à 12 mois. La ligne suivante indique le nombre de patients de l'échantillon d'apprentissage correspondant à cette feuille.

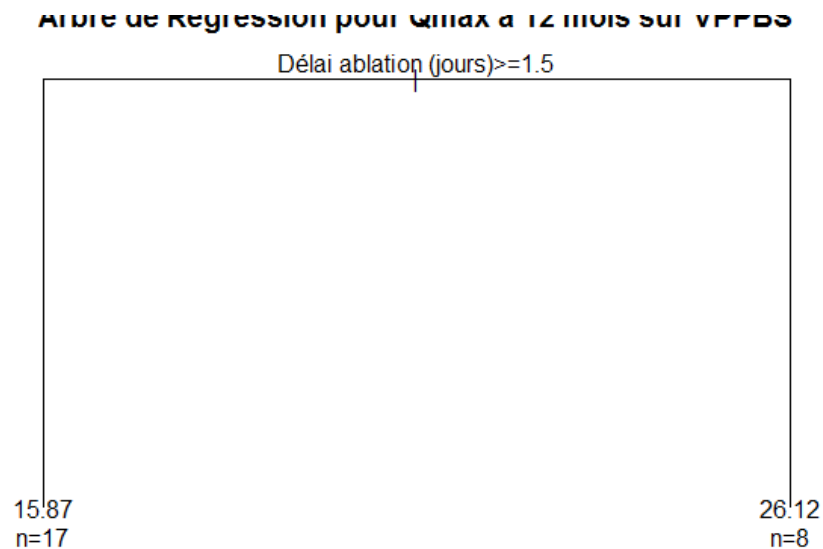


FIGURE 81 – VPPBS : ARBRE DE RÉGRESSION POUR QMAX À 12 MOIS

Pour évaluer l'arbre de régression, nous utiliserons une mesure de distance (valeur absolue de la différence) pour entre la valeur prédite et la valeur de référence pour chaque patient de l'ensemble de validation. La figure ?? représente l'écart constaté entre valeurs prédites (colonne de gauche) et valeurs de référence (colonne de droite). La moyenne de ces écarts nous donne une mesure du taux d'erreur de l'arbre. En répétant l'opération pour construire une forêt d'arbres de régression, nous utiliserons alors la prédiction rendue par l'arbre avec le plus faible taux d'erreur (Cf. figure ??).

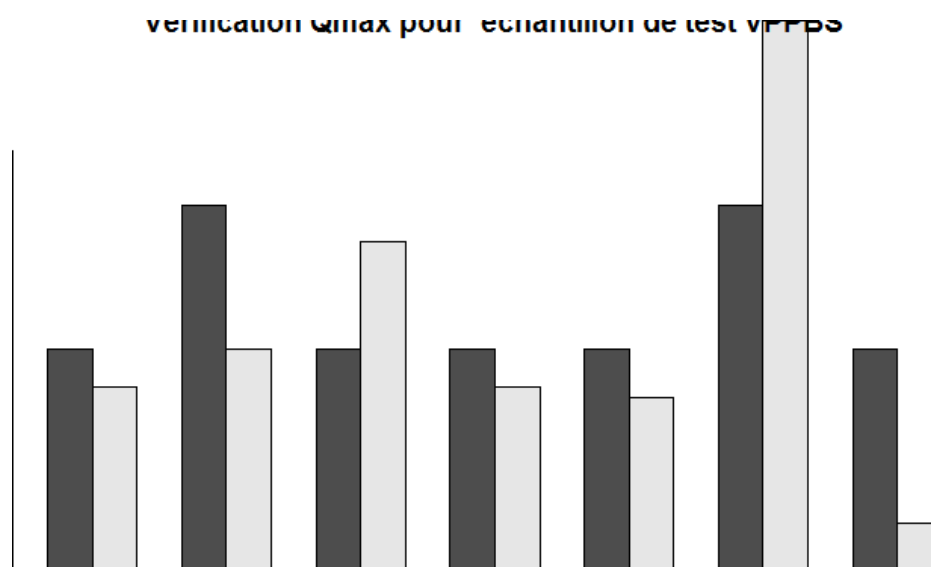


FIGURE 82 – VPPBS : EVALUATION DES PRÉVISIONS POUR QMAX À 12 MOIS

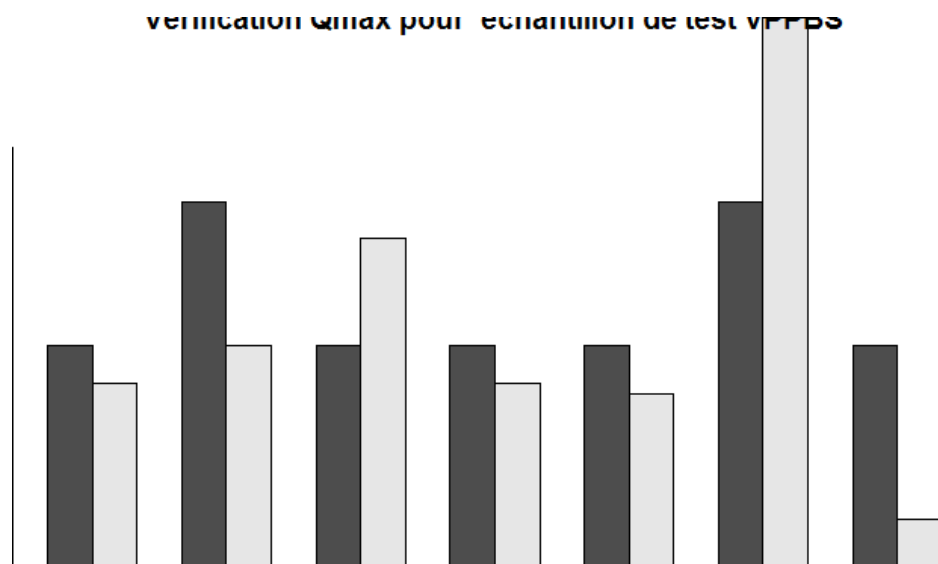


FIGURE 83 – VPPBS : EVALUATION DES PRÉVISIONS POUR QMAX À 12 MOIS

5.3 VAPOR : PRÉDICTION IPSS, QOL ET QMAX À 12 MOIS

5.3.1 VAPOR : IPSS À 12 MOIS

L'arbre de régression présenté figure ?? infère la valeur la plus probable pour IPSS à 12 mois à partir d'une seule variable, le volume prostatique et indique la probabilité de chaque valeur constatée dans l'ensemble d'apprentissage, respectivement 1/2/3/4/6/14.

Arbre de Régression pour IPSS à 12 mois sur VAPOR

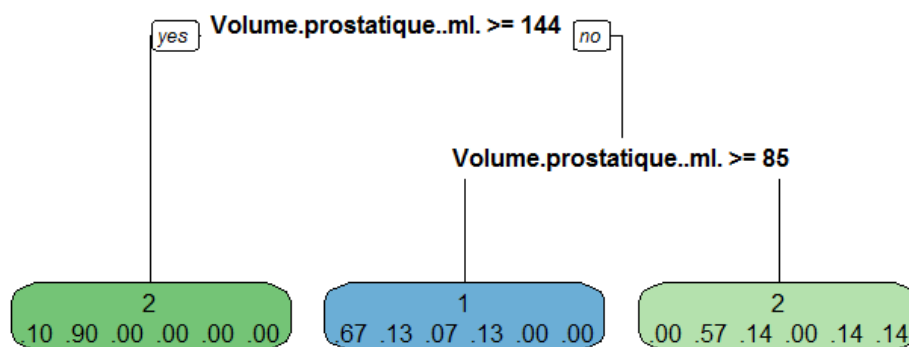


FIGURE 84 – VAPOR : ARBRE DE RÉGRESSION POUR IPSS À 12 MOIS

En appliquant cet arbre de régression sur l'ensemble de validation, nous obtenons une prédiction et une probabilité des valeurs IPSS, illustrées figure ??.

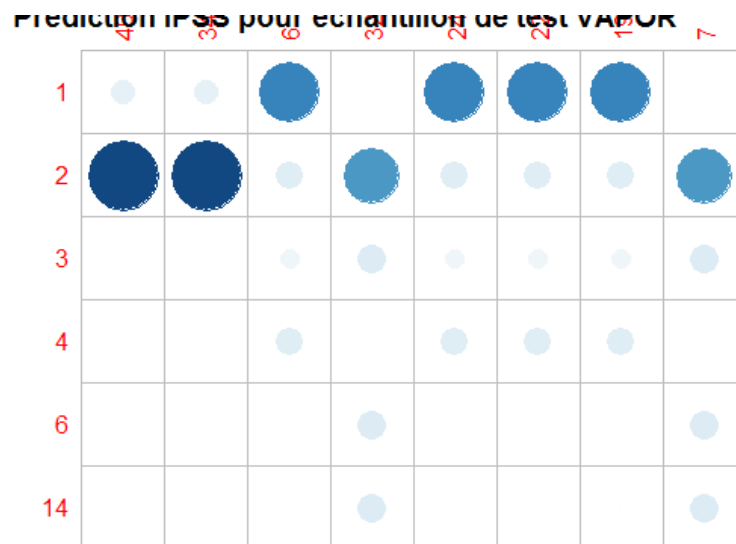


FIGURE 85 – VAPOR : PRÉVISION POUR IPSS À 12 MOIS

La table de correspondance suivant nous permet d'évaluer ces résultats :

prediction							
validation		1	2	3	4	6	14
	1	1	1	0	0	0	0
	2	1	2	0	0	0	0
	3	1	0	0	0	0	0
	4	1	0	0	0	0	0
	14	0	1	0	0	0	0

Soit un taux d'erreur de 62,5%.

En utilisant une forêt aléatoire, plutôt qu'un unique arbre de régression, nous obtenons alors une prédiction avec un taux d'erreur moindre, comme le montre cette nouvelle table de correspondance :

prediction							
validation		1	2	3	4	6	14
	1	2	0	0	0	0	0
	2	2	1	0	0	0	0
	3	1	0	0	0	0	0
	4	0	0	0	1	0	0
	14	0	0	0	0	1	0

Soit un taux d'erreur de 50%.

5.3.2 VAPOR : QOL À 12 MOIS

Arbre de Régression pour QOL à 12 mois sur VAPOR



FIGURE 86 – VAPOR : ARBRE DE RÉGRESSION POUR QOL À 12 MOIS

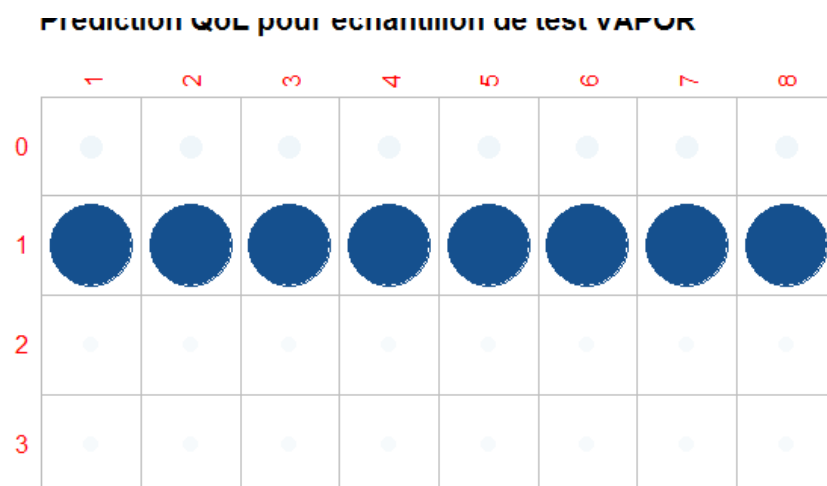


FIGURE 87 – VAPOR : PRÉVISION POUR QOL À 12 MOIS

	prediction				
validation	0	1	2	3	
	0	0	1	0	0
	1	0	6	0	0
	3	0	1	0	0

Soit un taux d'erreur de 25%.

Afin d'améliorer ce taux d'erreur de prédiction, Nous utilisons une forêt aléatoire d'arbres de régression. Nous obtenons alors une nouvelle table de correspondance :

	prediction			
validation	0	1	2	3
	0	0	1	0
	0	0	1	0
	1	0	6	0
	0	0	0	1

Soit un taux d'erreur de 12,5%.

5.3.3 VAPOR : QMAX À 12 MOIS

L'arbre de régression obtenu pour la variable Qmax à 12 mois (colonne Qmax (ml/s)_3) met en œuvre une seule variable (Cf. figure ??). Les feuilles de l'arbre représentent, en première ligne, la valeur approximative pour Qmax à 12 mois. La ligne suivante indique le nombre de patients de l'échantillon d'apprentissage correspondant à cette feuille.

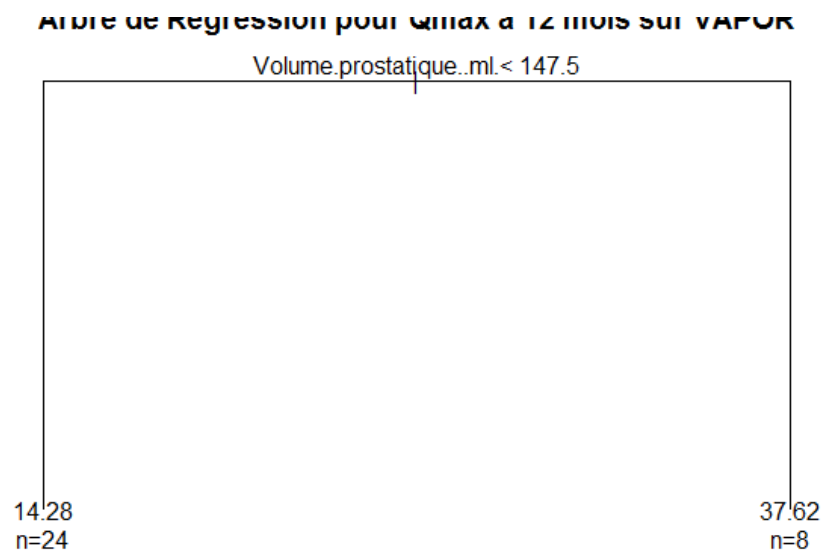


FIGURE 88 – VAPOR : ARBRE DE RÉGRESSION POUR QMAX À 12 MOIS

La figure ?? représente l'écart constaté entre valeurs prédites (colonne de gauche) et valeurs de référence (colonne de droite) sur l'échantillon de validation. La moyenne de ces écarts nous donne une mesure du taux d'erreur de l'arbre. En utilisant une forêt aléatoire d'arbres, figure ??, cet écart est réduit.

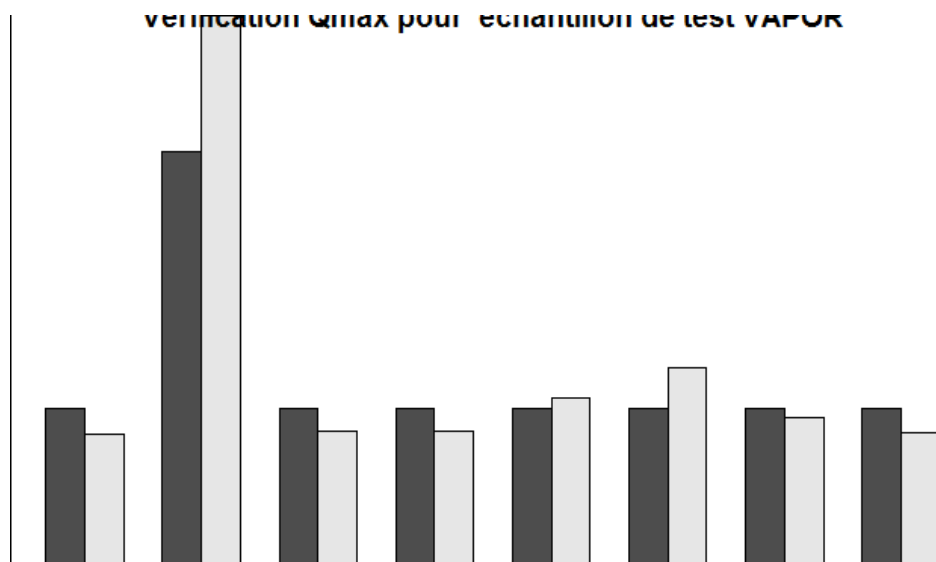


FIGURE 89 – VAPOR : EVALUATION DES PRÉVISIONS POUR QMAX À 12 MOIS

Vérification Qmax (RandomForest) pour échantillon de test VAPOR

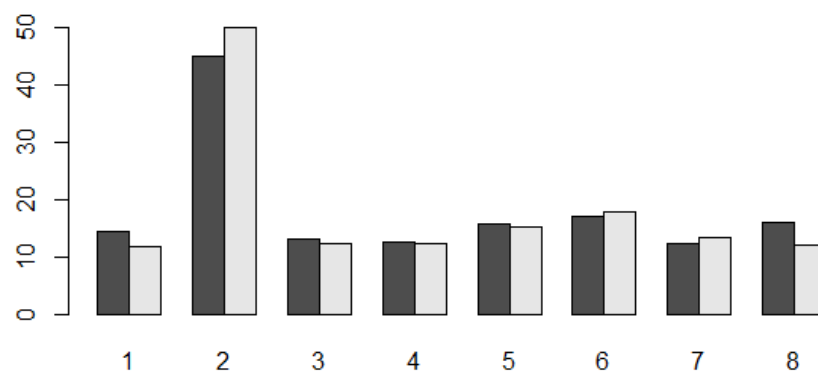


FIGURE 90 – VAPOR : EVALUATION DES PRÉVISIONS (RANDOM FOREST) POUR QMAX À 12 MOIS

5.4 RTUPB : PRÉDICTION DU PROFIL DE GUÉRISON

Nous utilisons ici la même méthode pour inférer arbres de régression et forêt aléatoire que lors de l'étude de prévision IPSS à 12 mois. A partir de l'arbre de régression complet obtenu sur l'échantillon d'apprentissage, la courbe évaluant le gain à chaque embranchement de l'arbre, figure ?? est quasi linéaire. Nous conservons donc cet arbre tel quel, sans élagage (figure ??).

FIGURE 91 – RTUPB : ARBRE DE RÉGRESSION POUR PROFIL DE GUÉRISON

Arbre de Régression pour classe de guérison sur RTUPB

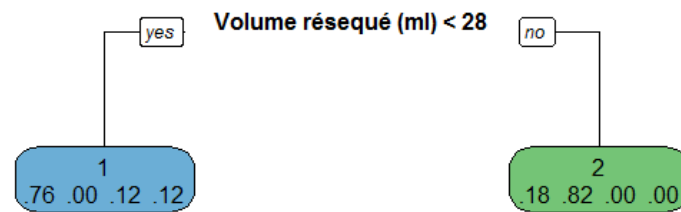


FIGURE 92 – RTUPB : ARBRE DE RÉGRESSION POUR PROFIL DE GUÉRISON

En appliquant l'arbre de régression sur l'échantillon de validation, nous obtenons des prévisions plutôt sûres d'elles (probabilité de 100%) comme illustré figure ??.

Prediction classe de guérison pour échantillon de test RTUPB

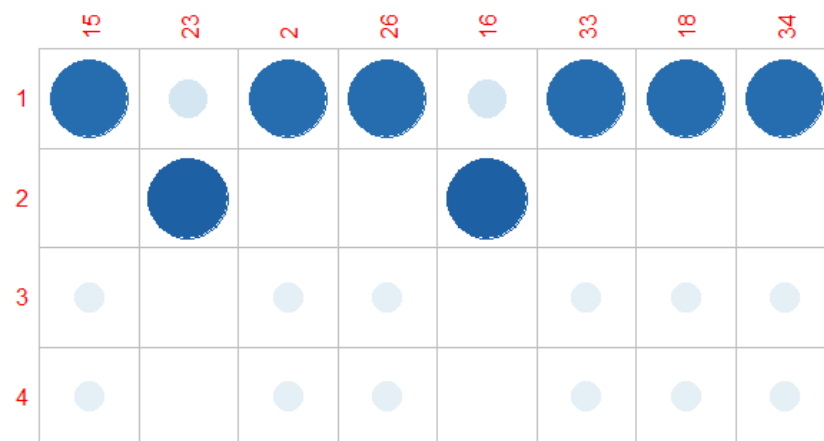


FIGURE 93 – RTUPB : PRÉDICTION DU PROFIL DE GUÉRISON

En comparant plus objectivement les prédictions avec les profils effectifs des patients de l'ensemble de validation, nous constatons, ci-dessous, un taux d'erreur de 12,5%.

prediction
validation 1 2 3 4

```

1 5 0 0 0
2 1 1 0 0
4 0 0 0 1

```

En utilisant une forêt aléatoire construite sur l'ensemble d'apprentissage, nous constatons alors, ci-dessous, un taux d'erreur de 0%. La classification des profils pré-opératoires, s'avère donc fiable pour recommander une indication d'opération RTUPB et le profil de guérison associé.

```

           prediction
validation 1 2 3 4
1 5 0 0 0
2 0 2 0 0
4 0 0 0 1

```

5.5 VPPBS : PRÉDICTION DU PROFIL DE GUÉRISON

Comme pour les exercices précédents, nous évaluons l'efficacité des embranchements de l'arbre de régression complet. La figure ?? montre un point d'inflexion pour $cp=0,13$. L'arbre élagué sur ce critère, est représenté figure ??.

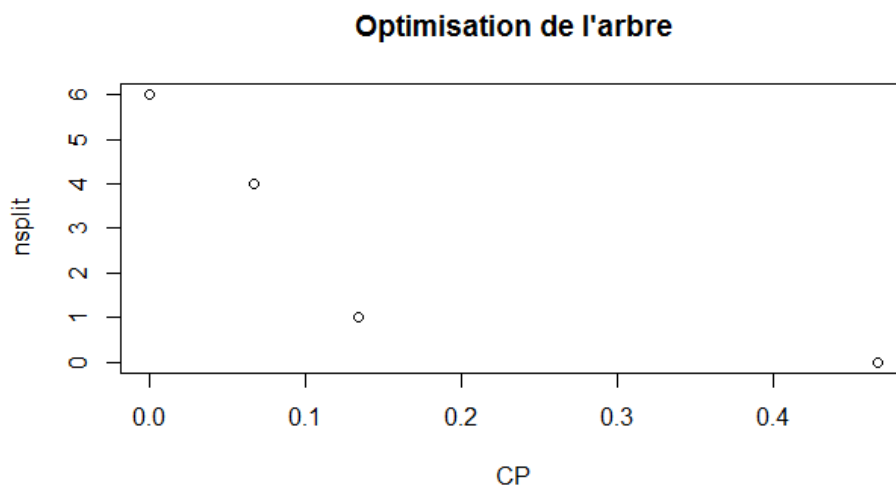


FIGURE 94 – VPPBS : OPTIMISATION DE L'ARBRE DE RÉGRESSION POUR PROFIL DE GUÉRISON

Arbre de Régression élagué pour classe de guérison sur VPPBS

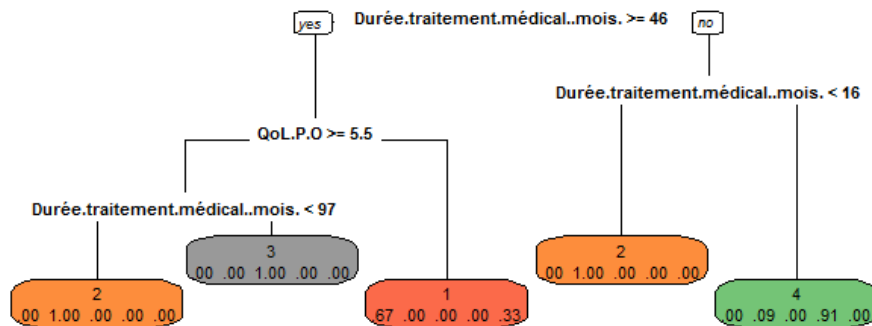


FIGURE 95 – VPPBS : ARBRE DE RÉGRESSION POUR PROFIL DE GUÉRISON

En appliquant l'arbre de régression sur l'échantillon de validation, nous obtenons des prévisions soit plutôt sûres (probabilités très fortes pour 3 patients sur 7), soit mitigées, comme illustré figure ??.

Prediction classe de guérison pour échantillon de test VPPBS

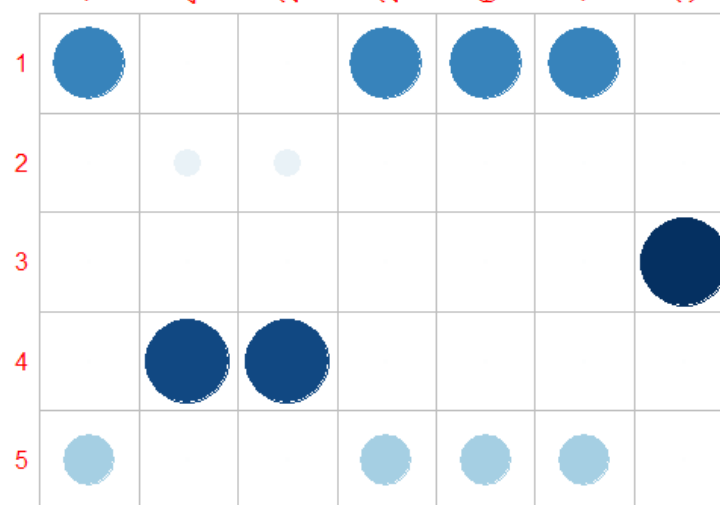


FIGURE 96 – VPPBS : PRÉDICTION DU PROFIL DE GUÉRISON

	prediction				
validation	1	2	3	4	5
1	1	0	0	0	0
2	0	0	0	1	0
3	0	0	1	0	0
4	0	0	0	1	0
5	3	0	0	0	0

Soit un taux d'erreur de 57%.

Par contre, en utilisant une forêt aléatoire construite sur l'ensemble d'apprentissage, nous constatons alors, ci-dessous, un taux d'erreur de 0%. La classification des profils pré-opérateurs, s'avère donc fiable pour recommander une indication d'opération VPPBS et le profil de guérison associé.

	prediction				
validation	1	2	3	4	5
1	1	0	0	0	0
2	0	1	0	0	0
3	0	0	1	0	0
4	0	0	0	1	0
5	0	0	0	0	3