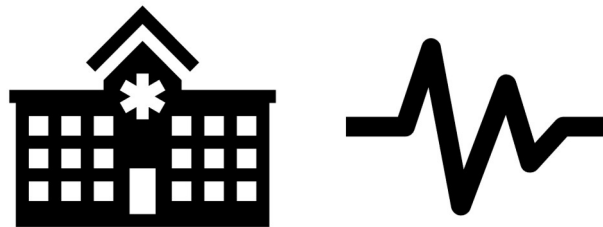


Etude de cas Biostatistique

Sujet : Comparaison des 2 groupes de prise en charge en termes de complications à 30 jours



Samba FALL – Lucas LE MAITRE – Hugo MARTINIERE

Mars – Avril 2023

Table des matières

1	Contexte	3
2	Notre jeu de données	3
3	Etudes comparatives de nos groupes	5
3.1	Variables qualitatives	5
3.2	Variable quantitative	6
4	Création de variables / formats	8
5	Odds – Radios	10
5.1	Analyses des P-values	11
5.2	Analyses des odds ratios	11
6	Création de modèles	12
6	Analyse de notre modèle final	13
6	Conclusion	14

1 Contexte

Cette étude statistique porte sur la comparaison des deux groupes de prise en charge, à savoir le groupe ambulatoire et l'hospitalisation conventionnelle, en termes de complications survenant dans les 30 jours suivant la prise en charge. Cette étude est menée dans un contexte où les pratiques médicales évoluent rapidement vers des modes de prise en charge plus courts et plus efficaces, notamment avec la montée en puissance des pratiques ambulatoires. Les résultats de cette étude pourront permettre de guider au mieux les professionnels de la santé quant au choix du mode de prise en charge le plus approprié tout en prenant en compte les risques de complications.

2 Notre jeu de données

Pour la réalisation de notre étude, nous disposons d'un jeu de données comprenant 34 variables. Ces variables fournissent des informations sur le patient, son traitement à l'hôpital, ainsi que sur d'éventuelles complications du patient après le traitement. Ci-dessous un tableau représentant les variables initialement présentes dans notre jeu de données avec leur libellé et les différentes valeurs possibles pour chaque variable.

Variables	Libellés	Valeurs
SUBJID	N° patient	Identifiant
DDN	Date de naissance	DATE
SEXE	Sexe	"HOMME";"FEMME"
DATE_INCLUSIOND1	Date d'inclusion	DATE
GROUPE_RANDO	Groupe randomisé (1 ou 2)	1=hospit conventionnelle, 2=ambulatoire
TABAC	Consommation de tabac	"oui <= 40 PA";"non ou sevré > 3 ans";"oui > 40 PA";"sevré <= 3 ans"
HTA	Hypertension artérielle	"NON";"OUI, traitée";"OUI, non traitée"
INSUF_RENALE	Insuffisance rénale	0;1
DYSLIPIDEMIE	Cholestérol et ou triglycérides	"OUI";"NON"
DIABETE	Diabète	0=NON 1=diabete de type1,2=diabete de type 2
IMC	Indice de masse corporelle	Nombre entre 17,3 et 35,9
CARDIO_ISCH_FAM	Antécédent familial de mort subite	"OUI";"NON"
MORT_SUBITE_FAM	Antécédent familial de cardiopathie ischémique	"OUI";"NON"
MALADIE_CEREVASC	Antécédent d'AVC	"OUI";"NON"
CARDIO_ISCH	Antécédent de cardiopathie ischémique	"OUI";"NON"
IC	Insuffisance cardiaque	"OUI";"NON"

ASA_INCLUSION	Score ASA	1;2;3
SCOREEVA_J0	Score état de santé de santé global	Score entre 20 et 100
RUTHERFORD_J0	Classification Rutherford à J0 (score de 0 à 5)	0;1;2;3;4;5;6
TTT_J0	Traitement à J0	"OUI";"NON"
DATE_ENTREE_HOSPITD1	Date d'entrée à l'hôpital	DATE
DATE_PROCEDED1	Date de la procédure	DATE
COTE_PONCTION	Coté de la ponction	"GAUCHE";"DROITE";"DROITE & GAUCHE"
ANESTH	Type d'anesthésie	"général";"locale+sedation";"loco régionale"
LONG_INTRO1	Longueur de l'introducteur	11;45;65
NBTOTALDEBALLONS	Nb total de ballon utilisé	0;1;2;3;4;5
NBTOTALDESTENTS	Nombre total de stent utilisé	0;1;2;3;4;5;7
antcdt_amput	Antécédent d'amputation	"OUI";"NON"
antcdt_revasc	antecedent de revascularisation	"OUI";"NON"
complic_pt_ponct	complication au point de ponction	"OUI";"NON"
complic_mineure	complication mineure	"OUI";"NON"
complicationM1		"OUI";"NON"
quoi	type de la complication	Caractère
AGE		Entier

3 Etudes comparatives de nos groupes

3.1 Variables qualitatives

Pour la comprendre le jeu de donnée sur lequel nous travaillons nous avons réalisé une analyse bivarié entre la variable Groupe Rando et les autres variables de notre jeu de donnée.

Ainsi l'objectif de cette partie est de trouver des variables influencé par le type de traitement.

Nous avons obtenu les résultats suivant pour toutes les variables qualitatives :

Var	Table	Statistic	Prob
1	Table SEXE * GROUPE_RANDO	Chi-Square	0.5590
2	Table TABAC * GROUPE_RANDO	Chi-Square	0.8036
3	Table HTA * GROUPE_RANDO	Chi-Square	0.2031
4	Table DIABETE * GROUPE_RANDO	Chi-Square	0.5244
5	Table CARDIO_ISCH_FAM * GROUPE_RANDO	Chi-Square	0.6139
6	Table MORT_SUBITE_FAM * GROUPE_RANDO	Chi-Square	0.4444
7	Table MALADIE_CEREVASC * GROUPE_RANDO	Chi-Square	0.3220
8	Table CARDIO_ISCH * GROUPE_RANDO	Chi-Square	0.3866
9	Table IC * GROUPE_RANDO	Chi-Square	0.3237
10	Table RUTHERFORD_J0 * GROUPE_RANDO	Chi-Square	0.6131
11	Table COTE_PONCTION * GROUPE_RANDO	Chi-Square	0.6898
12	Table ANESTH * GROUPE_RANDO	Chi-Square	0.5807
13	Table antcdt_amput * GROUPE_RANDO	Chi-Square	0.0763
14	Table antcdt_revasc * GROUPE_RANDO	Chi-Square	0.4232
15	Table complic_pt_ponct * GROUPE_RANDO	Chi-Square	0.3448
16	Table complic_mineure * GROUPE_RANDO	Chi-Square	0.1613
17	Table complicationM1 * GROUPE_RANDO	Chi-Square	0.1788

Puisque l'étude consistait à distinguer le lien entre les variables qualitatives on a donc utilisé le test de khi deux avec comme variable intérêt le groupe rando et pour variables explicatives les autres variables de notre base avec comme hypothèse :

- H_0 : les variables ont indépendantes si p
- H_1 : les variables sont dépendantes si p

A partir de notre tableau nous pouvons tirer comme première conclusion qu'il n'y a pas de différence significative entre les deux groupes. En effet, aucune P-Value de nos tests de khi-deux sur l'ensemble de ces variables n'a été concluant.

3.2 Variable quantitative

La seule variable quantitative présente est l'âge (avant recodage) puisque toutes les autres variables quantitatives ont été préalablement recodées avec des indicatrices

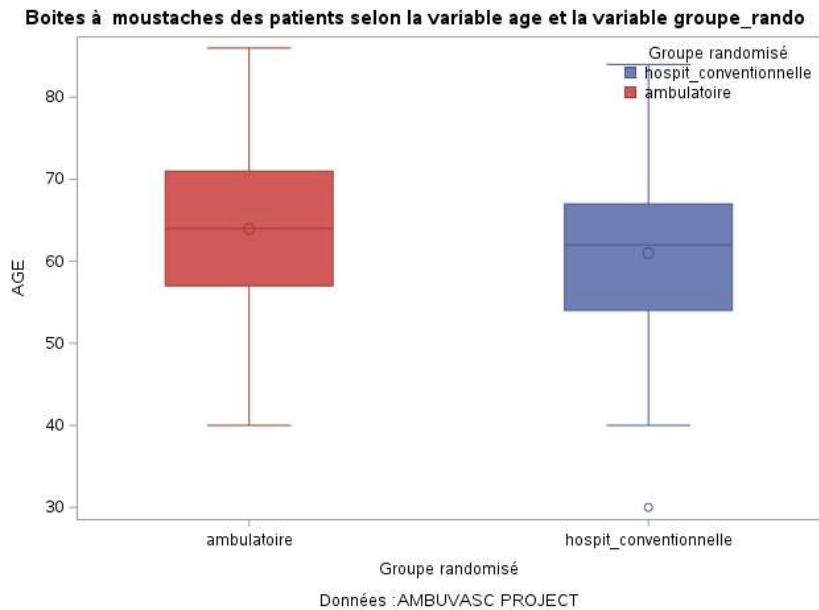
Notre seule variable quantitative présente dans notre jeu de donnée est la variable Age. Nous avons voulu tester sa corrélation avec la variable groupe pour cela nous avons préalablement fait le test de Shapiro Milk avec comme hypothèse :

- H_0 : l'échantillon suit une loi normale si la p_value
- H_1 : l'échantillon ne suit pas une loi normale si la p_value

Pour ce test nous avons trouvé une p-value de 0.0015 on en conclut que l'âge ne suit pas une loi normale. Ensuite, nous avons réalisé le test de Man Whitney pour définir si l'Age à un rôle significatif avec le type d'Hospitalisation avec comme hypothèse :

- H_0 Il n'y a pas de différence significative entre les moyennes des deux classes si p_value
- H_1 Y'a une différence significative entre les moyennes des deux classes si p_value

Le test de Man Whitney sur notre variable âge nous donne une p-value de 0.084 ce qui nous permet de conclure qu'il n'avait pas de différence significative entre les moyennes des patients dans le groupe ambulatoire et celui d'une l'hospitalisation conventionnel.



Cette boîte à moustache des âges en fonction du type d'Hospitalisation nous montre que la moyenne ainsi que la médiane d'âge des patients prises en charge par traitement ambulatoire est légèrement supérieur à celle des patients en hospit conventionnel écart (<2) ainsi ce résultat confirme nos tests effectués dans le paragraphe précédent (Il n'y a pas de différence significative).

Après cette première parti nous pouvons affirmer qu'aucune de nos variables n'est influencée par le type de traitement.

4 Création de variables / formats

Pour notre étude, nous avons dû transformer les variables quantitatives en variables qualitatives en recodant plusieurs d'entre elles. Par exemple, pour la variable "âge", que nous avons initialement obtenue avec la formule suivante : année d'entrée à l'hôpital du patient - année de naissance du patient.

Pour recoder cette variable, nous avons d'abord créé un tableau indiquant le pourcentage de chaque âge dans notre ensemble de données. Ensuite, nous avons sélectionné des tranches d'âge qui permettaient d'avoir le même pourcentage de patients dans chaque classe. Les classes d'âge que nous avons finalement retenues sont les suivantes : [35-57], [58-67], [67 et plus]. Pour l'affichage de nos visualisations, nous avons également créé des formats pour faciliter la compréhension de nos nouvelles variables.

Ci-dessous un tableau représentant les variables recodées avec leur Format associés et les différentes valeurs possibles pour chaque variable.

Nouvelles variables		Valeurs	Formats
new_antcdt_amput	0		NON
	1		OUI
new_antcdt_revasc	0		NON
	1		OUI
new_complic_pt_ponct	0		NON
	1		OUI
new_complic_mineure	0		NON
	1		OUI
new_DYSLIPIDEMIE	0		NON
	1		OUI
new_CARDIO_ISCH_FAM	0		NON
	1		OUI
new_MALADIE_CEREVASC	0		NON
	1		OUI
new_CARDIO_ISCH	0		NON
	1		OUI
new_IC	0		NON
	1		OUI

new_TTT_J0	0	NON
	1	OUI
new_IMC	Moins de 18,5 =1	Insuffisance pondérale (maigreur)
	18,5 à 25=2	Corpulence normale
	25 à 30=3	Surpoids
	30 à 35=4	Obésité modérée
	35 à 40=5	Obésité sévère
	Plus de 40=6	Obésité morbide ou massive
new_SCOREEVA_J0	[20;50[=1	Mauvaise santé
	[50;75]=2	Santé moyenne
]75;100]=3	Bonne santé
AGE	l'année de l'entrée en hospitalisation - l'année de naissance	AGE
new_rutherford	« 0,1,2 »=1	Santé convenable
	« 3,4,5,6 »=2	Mauvaise santé
new_age	Si AGE <57 =1	35-57
	Si AGE entre 58 et 67 =2	58-67
	Si AGE > 67 =3	67 et plus

5 Odds – Ratios

Ci-dessous le tableau de nos Odds – Ratio sur la variable complication sous 30 jours

Variable	Odds-ratio	OR[IC95%]	P-value global
new_antcdt_revasc NON vs YES	3.163	[1.127;8.881]	0,0288
new_antcdt_amput NON vs YES	17.865	[1.528;208.896]	0,0216
HTA NON vs OUI, traitée	0.307	[0.105;0.898]	0,0575
HTA OUI, non traitée vs OUI, traitée	0,136	[0.011;1.727]	0,0575
new_rutherford Mauvaise santé vs Santé convenable	1.741	[0.632;4.794]	0,2836
new_complic_pt_ponct NON vs YES	5.591	[1.756;17.799]	0,0036
IC NON vs OUI	8.867	[1.163;67.604]	0,035
GROUPE_RANDO ambulatoire vs hospit_conventionnelle	0.492	[0.172;1.406]	0,185
TABAC non ou sevré > 3 ans vs sevré <= 3 ans	0.355	[0.040;3.123]	0,56
TABAC oui <= 40 PA vs sevré <= 3 ans	0.230	[0.027;1.991]	0,56
TABAC oui > 40 PA vs sevré <= 3 ans	0.377	[0.037;3.879]	0,56
complic_mineure NON vs OUI	0.641	[0.078;5.260]	0,67
SEXE FEMME vs HOMME	999.999	[<0.001; >999.999]	0,97
new_age 35-57 vs 67 et plus	1.400	[0.413;4.741]	0,81
new_age 58-67 vs 67 et plus	1.400	[0.413;4.741]	0,81
new_MORT_SUBITE_FAM	0.003	[<0.001; >999.999]	0,978
new_MALADIE_CEREVASC	0.992	[0.116;8.457]	0,99
COTE_PONCTION Droit vs Gauche	1.140	[0.255;5.099]	0,38
COTE_PONCTION Droit & gauche vs Gauche	0.533	[0.134;2.128]	0,38
DIABETE NON vs diabete_de_type_2	1.179	[0.353;3.940]	0,96
DIABETE diabete_de_type_1 vs diabete_de_type_2	1.143	[0.111;11.722]	0,96
new_IMC Corpulence normale vs Surpoids	2.413	[0..759;7.671]	0,69
new_IMC Insuffisance pondérale vs Surpoids	>999.999	[<0.001; >999.999]	0,69
new_IMC Obésité modérée vs Surpoids	1.277	[0.312;5.216]	0,69
new_IMC Obésité sévère vs Surpoids	>999.999	[<0.001; >999.999]	0,69
new_SCOREEVA_J0 Bonne santé vs Santé moyenne	0.960	[0.301;3.058]	0,9518
new_SCOREEVA_J0 Mauvaise santé vs Santé moyenne	0.800	[0.198;3.233]	0,9519
CARDIO_ISCH NON vs OUI	1.523	[0.496;4.673]	0,4621
NBTOTALDESTENTS 0 vs 7	<0.001	[<0.001;>999.999]	0,7515
NBTOTALDESTENTS 1 vs 7	<0.001	[<0.001;>999.999]	0,7515
NBTOTALDESTENTS 2 vs 7	<0.001	[<0.001;>999.999]	0,7515
NBTOTALDESTENTS 3 vs 7	<0.001	[<0.001;>999.999]	0,7515
NBTOTALDESTENTS 4 vs 7	1.000	[<0.001;>999.999]	0,7515
NBTOTALDESTENTS 5 vs 7	<0.001	[<0.001;>999.999]	0,7515
NBTOTALDEBALLONS 0 vs 5	<0.001	[<0.001 ;>999.999]	0,5571
NBTOTALDEBALLONS 1 vs 5	<0.001	[<0.001 ;>999.999]	0,5571
NBTOTALDEBALLONS 2 vs 5	<0.001	[<0.001 ;>999.999]	0,5571
NBTOTALDEBALLONS 3 vs 5	<0.001	[<0.001 ;>999.999]	0,5571
NBTOTALDEBALLONS 4 vs 5	1.000	[<0.001 ;>999.999]	0,5571
LONG_INTRO1 11 vs 65	7.778	[0.447;135.457]	0,3216
LONG_INTRO1 45 vs 65	9.143	[0.513;162.798]	0,3216
ASA_INCLUSION 1 vs 3	1.037	[0.112;9.577]	0,8333
ASA_INCLUSION 2 vs 3	1.370	[0.484;3.880]	0,8333

Le tableau est le résultat de nos différents tests de corrélation entre les variables de notre base de données et notre variable à prédire « ComplicationM1 » ainsi on y trouve pour chaque variable sa p-value global et ses odds ratio associés.

5.1 Analyses des P-values

Pour cette partie l'objectif était de conserver nos variables qui avait des p-values < 0.2 , ce chiffre indiquant les variables expliquant le mieux la variable à prédire ainsi on a sélectionné toutes nos variables qui répondaient à ce critère plus la variable groupe_rando.

5.2 Analyses des odds ratios

L'odds ratio étant une mesure de probabilité qui représente le rapport des chances de succès (événement intérêt complicationm1) par rapport aux chances d'échecs. L'analyse était plus focalisée sur les variables antcsd-revasc antcdt_amput car ces variables étaient retenues pour le modèle finale ainsi pour Antcdt_amput on remarque que les personnes n'ayant pas d'antécédent amput avaient 17 fois plus de chances 'avoir des complications que ceux sans antécédents d'amputation. Ce résultats est à prendre avec des pincettes puisque seulement 3 individus sont concernés par ces antécédents ce qui est trop peu pour en tirer une conclusion.

En outre, nous pouvons dire qu'une personne arrivant en mauvaise santé à l'Hospital selon la classification de Rutherford a 1,741 fois plus de chance d'avoir des complications sous 30 jours qu'une personne ayant une santé convenable. On encore, une personne ayant une hypertension artérielle traitée à 3 fois de chance d'avoir des complications qu'une personne n'en ayant pas.

6-Création de modèles

Pour la création de notre modèles nous avons utilisé la méthode « Backward », nous avons dans un premier temps testé un modèle avec l'ensemble des variables présélectionnés, puis retirons chaque étape la variable comportant la P-value le plus élevé. Nous arrêtons ce processus lorsque toutes les P-Value sont en dessous de 5%. Nous avons alors fini par trouver le modèle suivant :

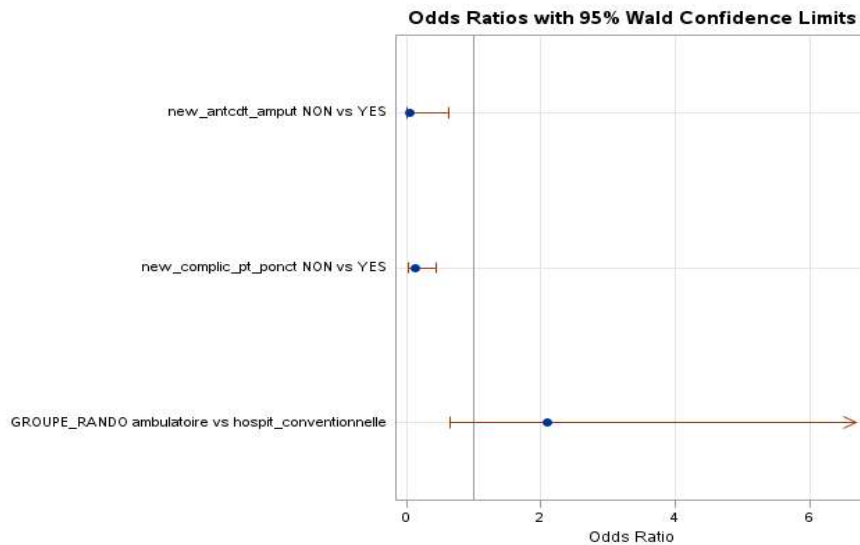
ComplicationM1 ; new_antcdt_amput, new_compli_pt_ponct, groupe_rando

La variable groupe_rando devant être gardé malgré une P-value > 5 puisqu'elle est essentielle pour notre.

Nous avons donc trouvé les résultats suivants pour notre modèle :

Variable	P-value
new_antcdt_amput	0,0206
new_compli_pt_ponct	0,0012
groupe_rando	0,2139

6-Analyse de notre modèle final



Dans cette dernière partie nous allons nous intéresser aux odds ratios, sur ce graphique plus les points bleus sont à droite plus la modalité de la variable exerce une influence significative sur la variable intérêt complicationM1 pour notre projet. De plus, si un point est également très proche de 0 alors sa variables associé est également très influente.

Ainsi on remarque que selon ce modèle les patients non touché par les variables new_antcdt_amput ainsi que new_antcdt_ponct ont beaucoup plus de chance d'avoir des complications, avec 20 fois plus pour les patients non touchés par des antécédents d'amputation ($OR=0,05$) et environ 7,5 fois plus pour les patients non touchés par une complication au point de ponction ($OR=0,131$). En outre, les personnes Hospitalisé de façon ambulatoire ont environ 2 fois plus de chance ($OR=2,089$) d'avoir des complications que ceux traité de façon conventionnelle avec un intervalle de confiance à 95% d'environ $[0,6 : 6,5]$.

Ces résultats sont à prendre avec précautions puisque comme rappelé précédemment la P-value entre le groupe randomisé et le lien entre les complications sous 30 jours était non significative. Ainsi que le faible nombre de patient impliqués dans certaine variable.

6 Conclusion

D'après les résultats de notre étude statistique, il semble qu'il n'y ait pas de lien entre une hospitalisation conventionnelle et une hospitalisation ambulatoire avec le risque de complications dans les 30 jours suivant la sortie de l'hôpital. Cela peut être une information importante pour les patients, les professionnels de la santé et les responsables politiques, car cela suggère que les deux types d'hospitalisation sont également sûrs en termes de risque de complications à court terme.