DEVOIR EPIDEMIOLOGIE : UTILISATION DU SCORE DE PROPENSION

Sarah F FELDMAN

27 février 2017

# INTRODUCTION

Le cathétérisme cardiaque droit (CCD) est un examen invasif utilisé en soins intensif pour mesurer directement la fonction cardiaque, ce qui permettrait selon certains médecins une meilleure prise en charge du patient et donc une augmentation de la survie. Cependant pour montrer une telle causalité il faudrait un essai randomisé. Or nous avons ici les données d'une cohorte prospective, non interventionnelle, de patients hostpitalisés en réanimation, certains ayant été cathétérisés, d'autres non. La cathétérisation n'a pas été randomisée, il y a donc lieu de penser que l'examen a été réalisé préférentiellement chez certains types de patients, potentiellement les patients les plus graves qui ont donc plus de risque de décéder. Cela constitue un biais d'indication qui rend l'analyse de l'efficacité du CCD ininterprêtable. Pour pouvoir analyser l'efficacité du CCD par sonde de swan Ganz, nous allons prendre en compte les biais d'indication propables à l'aide d'un score de propension, nous permettant d'améliorer le niveau de causalité de la relation CCD/Décès si elle existe.

# DESCRIPTION DE LA BASE DE DONNEES

[1] 5735 63

[1] 5735

Base de données de 5735 lignes et 63 colonnes. Ce sont des données transversales avec une ligne par patient (pas de doublon), chaque variable ayant été mesurées une seule fois.

# doublons

## Data management

[1] "CAT2"

FALSE   
 5735

FALSE TRUE   
 3719 3

0 1  
 FALSE 2005 1804  
 TRUE 8 1918

0 1   
3809 1918

0 1   
3817 1918

0% 25% 50% 75% 100%   
 2 7 12 22 338

0% 25% 50% 75% 100%   
 2 8 16 31 342

3 patients pour lesquels la date de décès est antérieure à la date de derniere nouvelle --> 2151 pour lesquels date de dernières nouvelles est antérieure au décès. -->

Avant d'utiliser la variable DTH30 dans une régression logistique, je vérifie que tous les sujets non décédés sont bien suivis au moins 30j. Si non, alors ils sont NA pour le décès et non pas 0 car on ne tient pas compte du temps. 8 sujets sont perdus de vu avant 30 jours et non décédés, ils sont donc NA pour le décès et non pas 0 comme noté dans la variable DTH30 (ils sont cependant censor = 0 pour une analyse de survie)

## Je réalise une analyse descriptive préliminaire afin de repérer les incohérences

### Présentation des variables de l'examen clinique et biologique:

Activities of daily living (ADL) : échelle d'autonomie de 0 à 12 (score>6 signe une dépendance)  
Duke Activity Status Index (DASI) : auto-questionnaire de 12 items mesurant l'activité fonctionnelle. Le score va de 0 à 58.2, plus le score est élevé, meilleur est l'activité fonctionnelle. DNR pour Do not resuscitate : 1 pour une interdiction de réanimation cardiopulmonaire, 0 sinon.  
Probabilité de survie à 2 mois, estimée par model : de 0 à 1  
APACHE III Acute Physiology scores : score de prediction du risque de mortalité de patients hospitalisés en soins intensifs. Plus le score est élevé plus le risque de mortalité est important. LE score va de 0 à 299 mais ici seule la partie physiologie du score est utilisée.  
Le score de Glasgow évalue l'état de conscience du patient: un score de 3 équivaut à un coma profond, un score de 15 est un état de conscience normal.  
Le rapport PAO2/FIO2 permet de diagnostiquer une agression pulmonaire aigue(rapport>300), un rapport <200 definissant le syndrome de detresse respiratoire aigue et l'ECMO est envisageable en cas de rapport <50. Un pH normal varie entre 7,38 et 7.42, le taux d'hématocrite normal est comprise entre 41 et 50% environ, La PaCO2 normale varie de 35 à 45mmHg.  
Le taux d'albumine normal varie entre 25 et 44 g/L.

[1] "ROWNAMES" "DEATH" "T3D30" "DTH30" "SWANG1" "PTID"   
 [7] "SWAN" "SWANT" "time" "DTH30b" "timeb"

Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique

valeur missing values  
SEX\_Female 2543 (44.3%)   
SEX\_Male 3192 (55.7%) 0 (0%)  
RACE\_black 920 (16%)   
RACE\_other 355 (6.2%) 0 (0%)  
RACE\_white 4460 (77.8%)   
INCOME\_$11-$25k 1165 (20.3%)   
INCOME\_$25-$50k 893 (15.6%) 0 (0%)  
INCOME\_> $50k 451 (7.9%)   
INCOME\_Under $11k 3226 (56.3%)   
NINSCLAS\_Medicaid 647 (11.3%)   
NINSCLAS\_Medicare 1458 (25.4%) 0 (0%)  
NINSCLAS\_Medicare & Medicaid 374 (6.5%)   
NINSCLAS\_No insurance 322 (5.6%)   
NINSCLAS\_Private 1698 (29.6%)   
NINSCLAS\_Private & Medicare 1236 (21.6%)   
AGE\* 64.05 (50.15-73.93) 0 (0%)  
EDU\* 12 (10-13) 0 (0%)  
CAT1\_ARF 2490 (43.4%)   
CAT1\_CHF 456 (8%) 0 (0%)  
CAT1\_Cirrhosis 224 (3.9%)   
CAT1\_Colon Cancer 7 (0.1%)   
CAT1\_Coma 436 (7.6%)   
CAT1\_COPD 457 (8%)   
CAT1\_Lung Cancer 39 (0.7%)   
CAT1\_MOSF w/Malignancy 399 (7%)   
CAT1\_MOSF w/Sepsis 1227 (21.4%)   
CAT2\_Cirrhosis 38 (0.7%)   
CAT2\_Colon Cancer 2 (0%) 0 (0%)  
CAT2\_Coma 90 (1.6%)   
CAT2\_Lung Cancer 15 (0.3%)   
CAT2\_MOSF w/Malignancy 229 (4%)   
CAT2\_MOSF w/Sepsis 826 (14.4%)   
CAT2\_NoCAT2 4535 (79.1%)   
CA\_No 4379 (76.4%)   
CA\_Metastatic 384 (6.7%) 0 (0%)  
CA\_Localized 972 (16.9%)   
CARDIOHX\_0 4722 (82.3%)   
CARDIOHX\_1 1013 (17.7%) 0 (0%)  
CHFHX\_0 4714 (82.2%)   
CHFHX\_1 1021 (17.8%) 0 (0%)  
DEMENTHX\_0 5171 (90.2%)   
DEMENTHX\_1 564 (9.8%) 0 (0%)  
PSYCHHX\_0 5349 (93.3%)   
PSYCHHX\_1 386 (6.7%) 0 (0%)  
CHRPULHX\_0 4646 (81%)   
CHRPULHX\_1 1089 (19%) 0 (0%)  
RENALHX\_0 5480 (95.6%)   
RENALHX\_1 255 (4.4%) 0 (0%)  
LIVERHX\_0 5334 (93%)   
LIVERHX\_1 401 (7%) 0 (0%)  
GIBLEDHX\_0 5550 (96.8%)   
GIBLEDHX\_1 185 (3.2%) 0 (0%)  
MALIGHX\_0 4419 (77.1%)   
MALIGHX\_1 1316 (22.9%) 0 (0%)  
IMMUNHX\_0 4192 (73.1%)   
IMMUNHX\_1 1543 (26.9%) 0 (0%)  
TRANSHX\_0 5073 (88.5%)   
TRANSHX\_1 662 (11.5%) 0 (0%)  
AMIHX\_0 5535 (96.5%)   
AMIHX\_1 200 (3.5%) 0 (0%)  
ADLD3P\* 0 (0-2) 4296 (74.9%)  
DAS2D3PC\* 19.75 (16.06-23.43) 0 (0%)  
DNR1\_0 5081 (88.6%)   
DNR1\_1 654 (11.4%) 0 (0%)  
SURV2MD1\* 0.63 (0.47-0.74) 0 (0%)  
APS1\* 54 (41-67) 0 (0%)  
SCOMA1\* 0 (0-41) 0 (0%)  
WTKILO1\* 70 (56.3-83.7) 0 (0%)  
TEMP1\* 38.09 (36.09-39) 0 (0%)  
MEANBP1\* 63 (50-115) 0 (0%)  
RESP1\* 30 (14-38) 0 (0%)  
HRT1\* 124 (97-141) 0 (0%)  
PAFI1\* 202.5 (133.31-316.62) 0 (0%)  
PACO21\* 37 (31-42) 0 (0%)  
PH1\* 7.4 (7.34-7.46) 0 (0%)  
WBLC1\* 14.1 (8.4-20.05) 0 (0%)  
HEMA1\* 30 (26.1-36.3) 0 (0%)  
SOD1\* 136 (132-142) 0 (0%)  
POT1\* 3.8 (3.4-4.6) 0 (0%)  
CREA1\* 1.5 (1-2.4) 0 (0%)  
BILI1\* 1.01 (0.8-1.4) 0 (0%)  
ALB1\* 3.5 (2.6-3.5) 0 (0%)  
URIN1\* 1927 (1110-2955) 3028 (52.8%)  
RESP\_0 3622 (63.2%)   
RESP\_1 2113 (36.8%) 0 (0%)  
CARD\_0 3804 (66.3%)   
CARD\_1 1931 (33.7%) 0 (0%)  
NEURO\_0 5042 (87.9%)   
NEURO\_1 693 (12.1%) 0 (0%)  
GASTR\_0 4793 (83.6%)   
GASTR\_1 942 (16.4%) 0 (0%)  
RENAL\_0 5440 (94.9%)   
RENAL\_1 295 (5.1%) 0 (0%)  
META\_0 5470 (95.4%)   
META\_1 265 (4.6%) 0 (0%)  
HEMA\_0 5381 (93.8%)   
HEMA\_1 354 (6.2%) 0 (0%)  
SEPS\_0 4704 (82%)   
SEPS\_1 1031 (18%) 0 (0%)  
TRAUMA\_0 5683 (99.1%)   
TRAUMA\_1 52 (0.9%) 0 (0%)  
ORTHO\_0 5728 (99.9%)   
ORTHO\_1 7 (0.1%) 0 (0%)  
 range  
SEX\_Female <NA>  
SEX\_Male <NA>  
RACE\_black <NA>  
RACE\_other <NA>  
RACE\_white <NA>  
INCOME\_$11-$25k <NA>  
INCOME\_$25-$50k <NA>  
INCOME\_> $50k <NA>  
INCOME\_Under $11k <NA>  
NINSCLAS\_Medicaid <NA>  
NINSCLAS\_Medicare <NA>  
NINSCLAS\_Medicare & Medicaid <NA>  
NINSCLAS\_No insurance <NA>  
NINSCLAS\_Private <NA>  
NINSCLAS\_Private & Medicare <NA>  
AGE\* 18.04199219 - 101.8479614  
EDU\* 0 - 30  
CAT1\_ARF <NA>  
CAT1\_CHF <NA>  
CAT1\_Cirrhosis <NA>  
CAT1\_Colon Cancer <NA>  
CAT1\_Coma <NA>  
CAT1\_COPD <NA>  
CAT1\_Lung Cancer <NA>  
CAT1\_MOSF w/Malignancy <NA>  
CAT1\_MOSF w/Sepsis <NA>  
CAT2\_Cirrhosis <NA>  
CAT2\_Colon Cancer <NA>  
CAT2\_Coma <NA>  
CAT2\_Lung Cancer <NA>  
CAT2\_MOSF w/Malignancy <NA>  
CAT2\_MOSF w/Sepsis <NA>  
CAT2\_NoCAT2 <NA>  
CA\_No <NA>  
CA\_Metastatic <NA>  
CA\_Localized <NA>  
CARDIOHX\_0 <NA>  
CARDIOHX\_1 <NA>  
CHFHX\_0 <NA>  
CHFHX\_1 <NA>  
DEMENTHX\_0 <NA>  
DEMENTHX\_1 <NA>  
PSYCHHX\_0 <NA>  
PSYCHHX\_1 <NA>  
CHRPULHX\_0 <NA>  
CHRPULHX\_1 <NA>  
RENALHX\_0 <NA>  
RENALHX\_1 <NA>  
LIVERHX\_0 <NA>  
LIVERHX\_1 <NA>  
GIBLEDHX\_0 <NA>  
GIBLEDHX\_1 <NA>  
MALIGHX\_0 <NA>  
MALIGHX\_1 <NA>  
IMMUNHX\_0 <NA>  
IMMUNHX\_1 <NA>  
TRANSHX\_0 <NA>  
TRANSHX\_1 <NA>  
AMIHX\_0 <NA>  
AMIHX\_1 <NA>  
ADLD3P\* 0 - 7  
DAS2D3PC\* 11 - 33  
DNR1\_0 <NA>  
DNR1\_1 <NA>  
SURV2MD1\* 0 - 0.961999893  
APS1\* 3 - 147  
SCOMA1\* 0 - 100  
WTKILO1\* 0 - 244  
TEMP1\* 27 - 43  
MEANBP1\* 0 - 259  
RESP1\* 0 - 100  
HRT1\* 0 - 250  
PAFI1\* 11.59960938 - 937.5  
PACO21\* 1 - 156  
PH1\* 6.579101563 - 7.76953125  
WBLC1\* 0 - 192  
HEMA1\* 2 - 66.1875  
SOD1\* 101 - 178  
POT1\* 1.099853516 - 11.8984375  
CREA1\* 0.099990845 - 25.09765625  
BILI1\* 0.099990845 - 58.1953125  
ALB1\* 0.299987793 - 29  
URIN1\* 0 - 9000  
RESP\_0 <NA>  
RESP\_1 <NA>  
CARD\_0 <NA>  
CARD\_1 <NA>  
NEURO\_0 <NA>  
NEURO\_1 <NA>  
GASTR\_0 <NA>  
GASTR\_1 <NA>  
RENAL\_0 <NA>  
RENAL\_1 <NA>  
META\_0 <NA>  
META\_1 <NA>  
HEMA\_0 <NA>  
HEMA\_1 <NA>  
SEPS\_0 <NA>  
SEPS\_1 <NA>  
TRAUMA\_0 <NA>  
TRAUMA\_1 <NA>  
ORTHO\_0 <NA>  
ORTHO\_1 <NA>

Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique

valeur missing values  
SEX\_Female 2543 (44.3%)   
SEX\_Male 3192 (55.7%) 0 (0%)  
RACE\_black 920 (16%)   
RACE\_other 355 (6.2%) 0 (0%)  
RACE\_white 4460 (77.8%)   
INCOME\_$11-$25k 1165 (20.3%)   
INCOME\_$25-$50k 893 (15.6%) 0 (0%)  
INCOME\_> $50k 451 (7.9%)   
INCOME\_Under $11k 3226 (56.3%)   
NINSCLAS\_Medicaid 647 (11.3%)   
NINSCLAS\_Medicare 1458 (25.4%) 0 (0%)  
NINSCLAS\_Medicare & Medicaid 374 (6.5%)   
NINSCLAS\_No insurance 322 (5.6%)   
NINSCLAS\_Private 1698 (29.6%)   
NINSCLAS\_Private & Medicare 1236 (21.6%)   
AGE\* 64.05 (50.15-73.93) 0 (0%)  
EDU\* 12 (10-13) 0 (0%)  
CAT1\_ARF 2490 (43.4%)   
CAT1\_CHF 456 (8%) 0 (0%)  
CAT1\_Cirrhosis 224 (3.9%)   
CAT1\_Colon Cancer 7 (0.1%)   
CAT1\_Coma 436 (7.6%)   
CAT1\_COPD 457 (8%)   
CAT1\_Lung Cancer 39 (0.7%)   
CAT1\_MOSF w/Malignancy 399 (7%)   
CAT1\_MOSF w/Sepsis 1227 (21.4%)   
CAT2\_Cirrhosis 38 (0.7%)   
CAT2\_Colon Cancer 2 (0%) 0 (0%)  
CAT2\_Coma 90 (1.6%)   
CAT2\_Lung Cancer 15 (0.3%)   
CAT2\_MOSF w/Malignancy 229 (4%)   
CAT2\_MOSF w/Sepsis 826 (14.4%)   
CAT2\_NoCAT2 4535 (79.1%)   
CA\_No 4379 (76.4%)   
CA\_Metastatic 384 (6.7%) 0 (0%)  
CA\_Localized 972 (16.9%)   
CARDIOHX\_0 4722 (82.3%)   
CARDIOHX\_1 1013 (17.7%) 0 (0%)  
CHFHX\_0 4714 (82.2%)   
CHFHX\_1 1021 (17.8%) 0 (0%)  
DEMENTHX\_0 5171 (90.2%)   
DEMENTHX\_1 564 (9.8%) 0 (0%)  
PSYCHHX\_0 5349 (93.3%)   
PSYCHHX\_1 386 (6.7%) 0 (0%)  
CHRPULHX\_0 4646 (81%)   
CHRPULHX\_1 1089 (19%) 0 (0%)  
RENALHX\_0 5480 (95.6%)   
RENALHX\_1 255 (4.4%) 0 (0%)  
LIVERHX\_0 5334 (93%)   
LIVERHX\_1 401 (7%) 0 (0%)  
GIBLEDHX\_0 5550 (96.8%)   
GIBLEDHX\_1 185 (3.2%) 0 (0%)  
MALIGHX\_0 4419 (77.1%)   
MALIGHX\_1 1316 (22.9%) 0 (0%)  
IMMUNHX\_0 4192 (73.1%)   
IMMUNHX\_1 1543 (26.9%) 0 (0%)  
TRANSHX\_0 5073 (88.5%)   
TRANSHX\_1 662 (11.5%) 0 (0%)  
AMIHX\_0 5535 (96.5%)   
AMIHX\_1 200 (3.5%) 0 (0%)  
ADLD3P\* 0 (0-2) 4296 (74.9%)  
DAS2D3PC\* 19.75 (16.06-23.43) 0 (0%)  
DNR1\_0 5081 (88.6%)   
DNR1\_1 654 (11.4%) 0 (0%)  
SURV2MD1\* 0.63 (0.47-0.74) 0 (0%)  
APS1\* 54 (41-67) 0 (0%)  
SCOMA1\* 0 (0-41) 0 (0%)  
WTKILO1\* 70 (56.3-83.7) 0 (0%)  
TEMP1\* 38.09 (36.09-39) 0 (0%)  
MEANBP1\* 63 (50-115) 0 (0%)  
RESP1\* 30 (14-38) 0 (0%)  
HRT1\* 124 (97-141) 0 (0%)  
PAFI1\* 202.5 (133.31-316.62) 0 (0%)  
PACO21\* 37 (31-42) 0 (0%)  
PH1\* 7.4 (7.34-7.46) 0 (0%)  
WBLC1\* 14.1 (8.4-20.05) 0 (0%)  
HEMA1\* 30 (26.1-36.3) 0 (0%)  
SOD1\* 136 (132-142) 0 (0%)  
POT1\* 3.8 (3.4-4.6) 0 (0%)  
CREA1\* 1.5 (1-2.4) 0 (0%)  
BILI1\* 1.01 (0.8-1.4) 0 (0%)  
ALB1\* 3.5 (2.6-3.5) 0 (0%)  
URIN1\* 1927 (1110-2955) 3028 (52.8%)  
RESP\_0 3622 (63.2%)   
RESP\_1 2113 (36.8%) 0 (0%)  
CARD\_0 3804 (66.3%)   
CARD\_1 1931 (33.7%) 0 (0%)  
NEURO\_0 5042 (87.9%)   
NEURO\_1 693 (12.1%) 0 (0%)  
GASTR\_0 4793 (83.6%)   
GASTR\_1 942 (16.4%) 0 (0%)  
RENAL\_0 5440 (94.9%)   
RENAL\_1 295 (5.1%) 0 (0%)  
META\_0 5470 (95.4%)   
META\_1 265 (4.6%) 0 (0%)  
HEMA\_0 5381 (93.8%)   
HEMA\_1 354 (6.2%) 0 (0%)  
SEPS\_0 4704 (82%)   
SEPS\_1 1031 (18%) 0 (0%)  
TRAUMA\_0 5683 (99.1%)   
TRAUMA\_1 52 (0.9%) 0 (0%)  
ORTHO\_0 5728 (99.9%)   
ORTHO\_1 7 (0.1%) 0 (0%)  
 range  
SEX\_Female <NA>  
SEX\_Male <NA>  
RACE\_black <NA>  
RACE\_other <NA>  
RACE\_white <NA>  
INCOME\_$11-$25k <NA>  
INCOME\_$25-$50k <NA>  
INCOME\_> $50k <NA>  
INCOME\_Under $11k <NA>  
NINSCLAS\_Medicaid <NA>  
NINSCLAS\_Medicare <NA>  
NINSCLAS\_Medicare & Medicaid <NA>  
NINSCLAS\_No insurance <NA>  
NINSCLAS\_Private <NA>  
NINSCLAS\_Private & Medicare <NA>  
AGE\* 18.04199219 - 101.8479614  
EDU\* 0 - 30  
CAT1\_ARF <NA>  
CAT1\_CHF <NA>  
CAT1\_Cirrhosis <NA>  
CAT1\_Colon Cancer <NA>  
CAT1\_Coma <NA>  
CAT1\_COPD <NA>  
CAT1\_Lung Cancer <NA>  
CAT1\_MOSF w/Malignancy <NA>  
CAT1\_MOSF w/Sepsis <NA>  
CAT2\_Cirrhosis <NA>  
CAT2\_Colon Cancer <NA>  
CAT2\_Coma <NA>  
CAT2\_Lung Cancer <NA>  
CAT2\_MOSF w/Malignancy <NA>  
CAT2\_MOSF w/Sepsis <NA>  
CAT2\_NoCAT2 <NA>  
CA\_No <NA>  
CA\_Metastatic <NA>  
CA\_Localized <NA>  
CARDIOHX\_0 <NA>  
CARDIOHX\_1 <NA>  
CHFHX\_0 <NA>  
CHFHX\_1 <NA>  
DEMENTHX\_0 <NA>  
DEMENTHX\_1 <NA>  
PSYCHHX\_0 <NA>  
PSYCHHX\_1 <NA>  
CHRPULHX\_0 <NA>  
CHRPULHX\_1 <NA>  
RENALHX\_0 <NA>  
RENALHX\_1 <NA>  
LIVERHX\_0 <NA>  
LIVERHX\_1 <NA>  
GIBLEDHX\_0 <NA>  
GIBLEDHX\_1 <NA>  
MALIGHX\_0 <NA>  
MALIGHX\_1 <NA>  
IMMUNHX\_0 <NA>  
IMMUNHX\_1 <NA>  
TRANSHX\_0 <NA>  
TRANSHX\_1 <NA>  
AMIHX\_0 <NA>  
AMIHX\_1 <NA>  
ADLD3P\* 0 - 7  
DAS2D3PC\* 11 - 33  
DNR1\_0 <NA>  
DNR1\_1 <NA>  
SURV2MD1\* 0 - 0.961999893  
APS1\* 3 - 147  
SCOMA1\* 0 - 100  
WTKILO1\* 0 - 244  
TEMP1\* 27 - 43  
MEANBP1\* 0 - 259  
RESP1\* 0 - 100  
HRT1\* 0 - 250  
PAFI1\* 11.59960938 - 937.5  
PACO21\* 1 - 156  
PH1\* 6.579101563 - 7.76953125  
WBLC1\* 0 - 192  
HEMA1\* 2 - 66.1875  
SOD1\* 101 - 178  
POT1\* 1.099853516 - 11.8984375  
CREA1\* 0.099990845 - 25.09765625  
BILI1\* 0.099990845 - 58.1953125  
ALB1\* 0.299987793 - 29  
URIN1\* 0 - 9000  
RESP\_0 <NA>  
RESP\_1 <NA>  
CARD\_0 <NA>  
CARD\_1 <NA>  
NEURO\_0 <NA>  
NEURO\_1 <NA>  
GASTR\_0 <NA>  
GASTR\_1 <NA>  
RENAL\_0 <NA>  
RENAL\_1 <NA>  
META\_0 <NA>  
META\_1 <NA>  
HEMA\_0 <NA>  
HEMA\_1 <NA>  
SEPS\_0 <NA>  
SEPS\_1 <NA>  
TRAUMA\_0 <NA>  
TRAUMA\_1 <NA>  
ORTHO\_0 <NA>  
ORTHO\_1 <NA>

[1] "1989-06-14" "1994-12-31"

[1] "1989-06-11" "1994-01-23"

0 1 2 3 4   
 269 2171 2711 583 1

ROWNAMES CAT1 CAT2 CA SADMDTE DSCHDTE DTHDTE LSTCTDTE   
 0 0 0 0 0 1 2013 0   
 DEATH CARDIOHX CHFHX DEMENTHX PSYCHHX CHRPULHX RENALHX LIVERHX   
 0 0 0 0 0 0 0 0   
GIBLEDHX MALIGHX IMMUNHX TRANSHX AMIHX AGE SEX EDU   
 0 0 0 0 0 0 0 0   
SURV2MD1 DAS2D3PC T3D30 DTH30 APS1 SCOMA1 MEANBP1 WBLC1   
 0 0 0 0 0 0 0 0   
 HRT1 RESP1 TEMP1 PAFI1 ALB1 HEMA1 BILI1 CREA1   
 0 0 0 0 0 0 0 0   
 SOD1 POT1 PACO21 PH1 SWANG1 WTKILO1 DNR1 NINSCLAS   
 0 0 0 0 0 0 0 0   
 RESP CARD NEURO GASTR RENAL META HEMA SEPS   
 0 0 0 0 0 0 0 0   
 TRAUMA ORTHO ADLD3P URIN1 RACE INCOME PTID SWAN   
 0 0 4296 3028 0 0 0 0   
 SWANT time DTH30b timeb   
 0 0 8 0

0 1 8 2013 3028 4296   
 63 1 1 1 1 1

FALSE TRUE   
0.2509154 0.7490846

FALSE TRUE   
0.4720139 0.5279861

[1] ROWNAMES CAT1 CAT2 CA SADMDTE DSCHDTE DTHDTE   
 [8] LSTCTDTE DEATH CARDIOHX CHFHX DEMENTHX PSYCHHX CHRPULHX  
[15] RENALHX LIVERHX GIBLEDHX MALIGHX IMMUNHX TRANSHX AMIHX   
[22] AGE SEX EDU SURV2MD1 DAS2D3PC T3D30 DTH30   
[29] APS1 SCOMA1 MEANBP1 WBLC1 HRT1 RESP1 TEMP1   
[36] PAFI1 ALB1 HEMA1 BILI1 CREA1 SOD1 POT1   
[43] PACO21 PH1 SWANG1 WTKILO1 DNR1 NINSCLAS RESP   
[50] CARD NEURO GASTR RENAL META HEMA SEPS   
[57] TRAUMA ORTHO ADLD3P URIN1 RACE INCOME PTID   
[64] SWAN SWANT time DTH30b timeb   
<0 rows> (or 0-length row.names)

ADLD3P et URIN1 ont respectivivement 75% et 53% de données manquantes tandis que les autres variables n'ont aucune ou moins de 5% de données manquantes. Je ne prendrai donc pas en compte ces deux variables dans le calcul du score de propension et je laisse les autres variables telles quelles pour le moment. Il n'y a pas de doublon.

## Je croise les variables pour chercher d'autres incohérences

0|0|0 0|0|TA0 0|FR0|0 0|FR0|TA0 FC0|0|0 FC0|0|TA0   
 5509 15 47 5 60 15   
 FC0|FR0|0 FC0|FR0|TA0   
 39 45

HRT1 RESP1 MEANBP1 DTH30 SCOMA1 TEMP1  
29 0 0 44 1 100 34.69531  
313 0 0 47 1 61 39.50000  
333 0 0 43 1 0 39.59375  
383 0 0 38 0 89 38.19531  
438 0 0 46 1 100 34.89844  
511 0 0 125 0 100 36.00000  
519 0 0 49 1 100 35.50000  
557 0 0 47 0 0 39.09375  
791 0 0 56 1 100 34.89844  
885 0 0 47 1 0 34.89844  
1207 0 0 116 0 41 38.79688  
1213 0 0 61 1 0 39.39844  
1232 0 0 29 1 26 38.79688  
1372 0 0 46 1 100 38.69531  
1430 0 0 41 1 100 38.89844  
1622 0 0 47 1 0 39.19531  
1708 0 0 129 0 0 40.09375  
1761 0 0 38 1 100 38.89844  
1809 0 0 45 1 100 37.00000  
1855 0 0 69 1 0 35.69531  
1871 0 0 63 0 94 38.79688  
1926 0 0 147 1 100 38.89844  
2710 0 0 146 0 0 39.29688  
2983 0 0 27 0 0 39.09375  
2984 0 0 51 1 100 36.00000  
3099 0 0 117 0 89 36.39844  
3318 0 0 47 1 100 41.29688  
3493 0 0 50 0 44 38.69531  
3679 0 0 162 1 55 40.69531  
3879 0 0 33 1 94 38.59375  
4239 0 0 118 1 9 39.69531  
4694 0 0 53 1 100 36.09375  
4782 0 0 210 1 100 39.79688  
5122 0 0 53 0 100 34.79688  
5202 0 0 63 1 100 35.29688  
5277 0 0 40 0 100 42.00000  
5415 0 0 33 0 100 34.89844  
5421 0 0 40 1 41 39.59375  
5431 0 0 55 0 0 35.00000

J'ai observé ensemble la fréquence cardiaque, la fréquence respiratoire et la tension artérielle. J'observe certaines incohérences : parfois 2 de ces variables peuvent valoir 0 mais pas la troisième, et parfois une seule de ces variables vaut 0.

0 1 2 3 4 5   
 250 1930 2694 796 64 1

ROWNAMES CAT1 CAT2 CA SADMDTE DSCHDTE DTHDTE LSTCTDTE   
 0 0 0 0 0 1 2013 0   
 DEATH CARDIOHX CHFHX DEMENTHX PSYCHHX CHRPULHX RENALHX LIVERHX   
 0 0 0 0 0 0 0 0   
GIBLEDHX MALIGHX IMMUNHX TRANSHX AMIHX AGE SEX EDU   
 0 0 0 0 0 0 0 0   
SURV2MD1 DAS2D3PC T3D30 DTH30 APS1 SCOMA1 MEANBP1 WBLC1   
 0 0 0 0 0 0 15 0   
 HRT1 RESP1 TEMP1 PAFI1 ALB1 HEMA1 BILI1 CREA1   
 60 0 0 0 2 8 0 0   
 SOD1 POT1 PACO21 PH1 SWANG1 WTKILO1 DNR1 NINSCLAS   
 0 0 21 0 0 515 0 0   
 RESP CARD NEURO GASTR RENAL META HEMA SEPS   
 0 0 0 0 0 0 0 0   
 TRAUMA ORTHO ADLD3P URIN1 RACE INCOME PTID SWAN   
 0 0 4296 3028 0 0 0 0   
 SWANT time DTH30b timeb   
 0 0 8 0

0 1 2 8 15 21 60 515 2013 3028 4296   
 57 1 1 2 1 1 1 1 1 1 1

Les scores qui ont été mesurés, semblent tous avoir des résultats plausibles excepté le score de Glasgow qui a été modifié : d'un score allant de 3 à 15 on passe à un score allant de 0 à 100. Le score de 100 est probablement lié à un coma sévère car les patients avec un score plus élevé ont un risque de décès augmenté (alors que dans l'échelle d'origine, 3 correspond au coma sévère, et 15 à une conscience normale).  
Je laisse la variable telle quelle car elle a de toute évidence été modifiée, je ne veux pas la remodifier une seconde fois.  
La température corporelle peut aller de 27° en cas d'hypothermie profonde à 42-43° en cas de forte fièvre, donc je ne modifie pas.  
Un patient ne peut pas avoir un poids de 0, ces 515 patients sont recodés en NA pour le poids.  
Un patient peut être en aplasie mais je ne suis pas sûr qu'une hyperleucocytose de plus de 40 leucocytes/10^9L soit possible, cependant dans le doute concernant le taux maximum en cas de leucémie et au vue de la distribution de la variable qui semble plausible, je laisse telle quelle.  
Hematocrite : une anémie chronique peut entrainer une anémie profonde avec une hématocrite de 10%, en dessous de cette valeur je mets NA. (un taux de 66% est possible). Un taux de PAcO2 de 156 mmHg me semble complètement impossible, au delà de 100 mmHg je met NA. Un pH de 6.6 est extrêmement faible et n'est pas viable mais je ne suis pas sûre qu'on ne puisse pas l'observer en réanimation.  
Une créatininémie de 2200umol/L ou 25 mg/dL est possible en cas d'insuffisance rénale terminale. Un volume urinaire de 0 et de 9000mL est possible également.  
Une valeur normale d'albumine se situe entre 3.5 et 5 g/dL. Les patients avec un taux d'albumine supérieur à 10g/dL auront une valeur NA pour l'albumine.

Les valeurs de natrémie, de kaliémie, de fréquence respiratoire, fréquence cardiaque, tension artéirelle et bilirubine semblent plausible

<http://www.srlf.org/wp-content/uploads/2015/11/0505-Reanimation-Vol14-N3-p177_185.pdf>

Règles de recodage de la fréquence caridaque, de la fréquence respiratoire et de la tension artérielle: - Si deux de ces variables valait 0, alors la valeur de 0 était systématiquement attribuée à la troisième.  
- Si une seule de ces variables valait 0 alors elle était systématiquement transformée en valeur manquante. Recodage du poids : si le poids est inférieur à 25 je remplace par une valeure manquante.

# Je regarde à quoi correspond la variable SCOMA1

# Possible erreur des variables T3D30 et DTH30

timeb time T3D30 DEATH DTH30  
453 4 4 30 0 0  
488 6 6 30 0 0  
2338 21 21 30 0 0  
2811 10 10 30 0 0  
2907 2 2 30 0 0  
4783 9 9 30 0 0  
5123 28 28 30 0 0  
5437 8 8 30 0 0

8 sujets sont notés comme ayant une durée de suivi de 30 jours, or ils sont en fait perdus de vu avant 30 jours. J'ai donc recré une variable durée de suivi, time, qui est la différence entre entre la date des dernières nouvelles (décès ou date de dernier contact) et la date d'admission, et j'ai ensuite crée la variable timeb telle que si cette durée était supérieure à 30, elle prenait alors la valeur 30. en comparant la variable timeb à la variable T3D30, je vois que les variables sont égales 2 à 2 pour chaque sujets exceptés les 8 sujets dont j'ai parlé ci dessus. J'ai donc remplacé la valeur de T3D30 pour ces sujets là par la valeure de timeb. Ces 8 sujets sont notés comme n'ayant pas eu d'évènement à 30 jours. Cela ne pose pas de problème si l'on fait un modèle de Cox qui prend en compte les perdus de vue, mais je souhaitais faire une régression logistique conditionnelle et dans ce cas je ne peux pas dire que les 8 sujets perdus de vue avant 30 jours ne sont pas mort à 30 jours et je choisi donc de mettre une donnée manquante pour ces 8 valeurs

# Description des variables

Je décris les variables dans le groupe des patients cathétérisés, dans le groupe des patients non cathétérisés et les deux groupes confondus.

Variables quantitatives: - continues : - discrètes : - binaires variables qualitatives

Pause dans les inclusions

# Description avant imputation

Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique

# Imputation des valeurs manquantes

Je décide de ne garder que les variables avec moins de 20%de données manquantes et les sujets avec moins de 50% de données manquantes (excepté la variable date de décès bien sûr). Je ne tiens compte que des variables explicatives (je ne tiens pas compte du décès ni du traitement, qui n'est jamais NA de toutes façons, ni des dates d'entrée dans l'étude ou de sortie de l'hopital). J'imputerai les données manquantes restantes.

Je vais donc dans un premier temps observer les données manquantes par variable et par sujet.

ADLD3P URIN1   
74.90846 52.79861

napat1  
 0 1 2 3 4   
 581 2565 2390 196 3

napat  
 0 1.47058823529412 2.94117647058824 4.41176470588235   
 250 1930 2694 796   
5.88235294117647 7.35294117647059   
 64 1

Le score d'autonomie ADL a environ 75% de données manquantes et la variable volume d'excrétion urinaire a plus de 50% de données manquantes. Je supprime donc ces deux variables de l'analyse. 250 patient patients n'ont aucune donnée manquante, 5485(95.6%) patients ont au moins 1 donnée manquante dont : 1930 patients avec 1 donnée manquante, 2694 patients avec 2 de données manquantes, 796 patients avec 3 données manquantes, 64 patients avec 4 données manquantes et 1 patient avec 5 données manquantes. Tous les patients ont donc moins de 50% de données manquantes. Je les garde tous dans l'analyse.

Je note également qu'il n'y a pas de valeurs manquantes concernant la variable d'intérêt Swan ganz. Si je prends comme outcome le décès à 30 jours, analysé par régression logistique, j'ai 8 valeurs manquantes (car les sujets ont été censurés avant 30j, et la censure n'est pas prise en charge par le modèle logistique).

MEANBP1 HRT1 ALB1 HEMA1 PACO21 WTKILO1   
0.26155187 1.04620750 0.03487358 0.13949433 0.36617262 8.97994769

napat1  
 0 1 2   
5122 605 8

napat  
 0 1.66666666666667 3.33333333333333   
 5122 605 8

En supprimant ces 2 colonnes de l'analyse, j'ai 2397 patients avec au moins une donnée manquante, soit plus de 5 % de patients. Je ne peux donc pas simplement supprimer les patients avec des données manquantes. 6 variables ont au moins une donnée manquante : la tension artérielle (15 NA), la fréquence cardiaque (60 NA), l'albuminémie (2 NA) l'hématocrite (8 NA), la PaCO2 (21 NA) et le poids (515 NA).

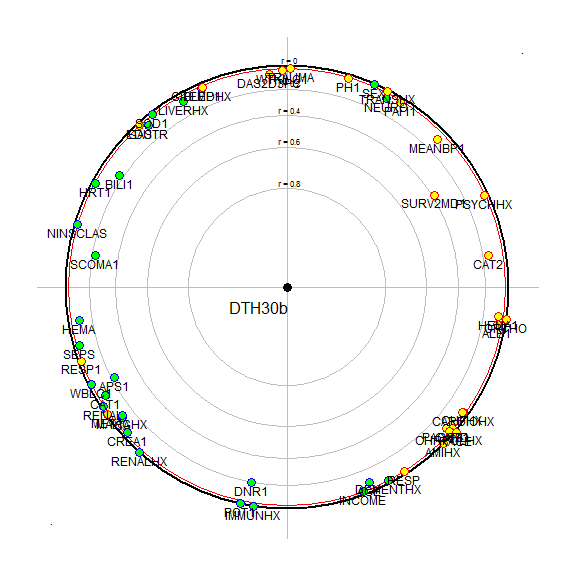
Dans un second temps, je réaliserai une imputation multiple en faisant l'hypothèse que les données manquantes le sont aléatoirement.

iter imp variable  
 1 1 MEANBP1 HRT1 ALB1 HEMA1 PACO21 WTKILO1 DTH30b  
 2 1 MEANBP1 HRT1 ALB1 HEMA1 PACO21 WTKILO1 DTH30b  
 3 1 MEANBP1 HRT1 ALB1 HEMA1 PACO21 WTKILO1 DTH30b  
 4 1 MEANBP1 HRT1 ALB1 HEMA1 PACO21 WTKILO1 DTH30b  
 5 1 MEANBP1 HRT1 ALB1 HEMA1 PACO21 WTKILO1 DTH30b

# Description après imputation

Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique  
  
Warning in FUN(X[[i]], ...): NAs introduits lors de la conversion  
automatique

# Exploratoire



# Tests bivariés

[1] "RESP"  
[1] "CARD"  
[1] "NEURO"  
[1] "GASTR"  
[1] "RENAL"  
[1] "META"  
[1] "HEMA"  
[1] "SEPS"  
[1] "TRAUMA"  
[1] "ORTHO"  
[1] "CAT1"  
[1] "CAT2"  
[1] "CA"  
[1] "DAS2D3PC"  
[1] "DNR1"  
[1] "SURV2MD1"  
[1] "APS1"  
[1] "SCOMA1"  
[1] "WTKILO1"  
[1] "TEMP1"  
[1] "MEANBP1"  
[1] "RESP1"  
[1] "HRT1"  
[1] "PAFI1"  
[1] "PACO21"  
[1] "PH1"  
[1] "WBLC1"  
[1] "HEMA1"  
[1] "SOD1"  
[1] "POT1"  
[1] "CREA1"  
[1] "BILI1"  
[1] "ALB1"  
[1] "CARDIOHX"  
[1] "CHFHX"  
[1] "DEMENTHX"  
[1] "PSYCHHX"  
[1] "CHRPULHX"  
[1] "RENALHX"  
[1] "LIVERHX"  
[1] "GIBLEDHX"  
[1] "MALIGHX"  
[1] "IMMUNHX"  
[1] "TRANSHX"  
[1] "AMIHX"  
[1] "AGE"  
[1] "SEX"  
[1] "RACE"  
[1] "EDU"  
[1] "INCOME"  
[1] "NINSCLAS"

[1] "RESP"  
[1] "CARD"  
[1] "NEURO"  
[1] "GASTR"  
[1] "RENAL"  
[1] "META"  
[1] "HEMA"  
[1] "SEPS"  
[1] "TRAUMA"  
[1] "ORTHO"  
[1] "CAT1"  
[1] "CAT2"  
[1] "CA"  
[1] "DAS2D3PC"  
[1] "DNR1"  
[1] "SURV2MD1"  
[1] "APS1"  
[1] "SCOMA1"  
[1] "WTKILO1"  
[1] "TEMP1"  
[1] "MEANBP1"  
[1] "RESP1"  
[1] "HRT1"  
[1] "PAFI1"  
[1] "PACO21"  
[1] "PH1"  
[1] "WBLC1"  
[1] "HEMA1"  
[1] "SOD1"  
[1] "POT1"  
[1] "CREA1"  
[1] "BILI1"  
[1] "ALB1"  
[1] "CARDIOHX"  
[1] "CHFHX"  
[1] "DEMENTHX"  
[1] "PSYCHHX"  
[1] "CHRPULHX"  
[1] "RENALHX"  
[1] "LIVERHX"  
[1] "GIBLEDHX"  
[1] "MALIGHX"  
[1] "IMMUNHX"  
[1] "TRANSHX"  
[1] "AMIHX"  
[1] "AGE"  
[1] "SEX"  
[1] "RACE"  
[1] "EDU"  
[1] "INCOME"  
[1] "NINSCLAS"

coef pvalue SWAN significatif SWAN coef pvalue DEATH  
RESP 0.000 \* 0.002  
CARD 0.000 \* 0.000  
NEURO 0.000 \* 0.000  
GASTR 0.000 \* 0.000  
HEMA 0.026 \* 0.000  
SEPS 0.000 \* 0.000  
CAT1 0.000 \* 0.000  
CAT2 0.000 \* 0.000  
CA 0.000 \* 0.000  
DAS2D3PC 0.023 \* 0.000  
DNR1 0.000 \* 0.000  
SURV2MD1 0.000 \* 0.000  
APS1 0.000 \* 0.000  
SCOMA1 0.000 \* 0.000  
WTKILO1 0.000 \* 0.002  
TEMP1 0.429 0.011  
MEANBP1 0.000 \* 0.000  
PAFI1 0.000 \* 0.001  
PACO21 0.000 \* 0.000  
PH1 0.000 \* 0.015  
WBLC1 0.002 \* 0.020  
HEMA1 0.000 \* 0.000  
CREA1 0.000 \* 0.000  
BILI1 0.000 \* 0.000  
ALB1 0.000 \* 0.000  
CARDIOHX 0.000 \* 0.000  
CHFHX 0.010 \* 0.000  
DEMENTHX 0.000 \* 0.001  
PSYCHHX 0.000 \* 0.001  
CHRPULHX 0.000 \* 0.000  
LIVERHX 0.075 0.000  
GIBLEDHX 0.012 \* 0.007  
MALIGHX 0.000 \* 0.000  
AGE 0.026 \* 0.000  
INCOME 0.000 \* 0.000  
NINSCLAS 0.000 \* 0.000  
 significatif DEATH select  
RESP \* 1  
CARD \* 1  
NEURO \* 1  
GASTR \* 1  
HEMA \* 1  
SEPS \* 1  
CAT1 \* 1  
CAT2 \* 1  
CA \* 1  
DAS2D3PC \* 1  
DNR1 \* 1  
SURV2MD1 \* 1  
APS1 \* 1  
SCOMA1 \* 1  
WTKILO1 \* 1  
TEMP1 \* 1  
MEANBP1 \* 1  
PAFI1 \* 1  
PACO21 \* 1  
PH1 \* 1  
WBLC1 \* 1  
HEMA1 \* 1  
CREA1 \* 1  
BILI1 \* 1  
ALB1 \* 1  
CARDIOHX \* 1  
CHFHX \* 1  
DEMENTHX \* 1  
PSYCHHX \* 1  
CHRPULHX \* 1  
LIVERHX \* 1  
GIBLEDHX \* 1  
MALIGHX \* 1  
AGE \* 1  
INCOME \* 1  
NINSCLAS \* 1

[1] 36

c("RESP", "CARD", "NEURO", "GASTR", "HEMA", "SEPS", "CAT1", "CAT2",   
"CA", "DAS2D3PC", "DNR1", "SURV2MD1", "APS1", "SCOMA1", "WTKILO1",   
"TEMP1", "MEANBP1", "PAFI1", "PACO21", "PH1", "WBLC1", "HEMA1",   
"CREA1", "BILI1", "ALB1", "CARDIOHX", "CHFHX", "DEMENTHX", "PSYCHHX",   
"CHRPULHX", "LIVERHX", "GIBLEDHX", "MALIGHX", "AGE", "INCOME",   
"NINSCLAS")

[1] "RESP" "CARD" "NEURO" "GASTR" "HEMA" "SEPS"   
 [7] "CAT1" "CAT2" "CA" "DAS2D3PC" "DNR1" "SURV2MD1"  
[13] "APS1" "SCOMA1" "WTKILO1" "TEMP1" "MEANBP1" "PAFI1"   
[19] "PACO21" "PH1" "WBLC1" "HEMA1" "CREA1" "BILI1"   
[25] "ALB1" "CARDIOHX" "CHFHX" "DEMENTHX" "PSYCHHX" "CHRPULHX"  
[31] "LIVERHX" "GIBLEDHX" "MALIGHX" "AGE" "INCOME" "NINSCLAS"

[1] "RESP" "CARD" "NEURO" "GASTR" "HEMA" "SEPS"   
 [7] "CAT1" "CAT2" "CA" "DAS2D3PC" "DNR1" "SURV2MD1"  
[13] "APS1" "SCOMA1" "WTKILO1" "MEANBP1" "PAFI1" "PACO21"   
[19] "PH1" "WBLC1" "HEMA1" "CREA1" "BILI1" "ALB1"   
[25] "CARDIOHX" "CHFHX" "DEMENTHX" "PSYCHHX" "CHRPULHX" "GIBLEDHX"  
[31] "MALIGHX" "AGE" "INCOME" "NINSCLAS"

[1] "TEMP1" "LIVERHX"

[1] "RENAL" "TRAUMA" "RESP1" "HRT1" "SOD1" "IMMUNHX" "TRANSHX"  
 [8] "AMIHX" "SEX" "EDU"

J'ai 9 classes au maximum (variable CAT1), donc la condition des 5 à 10 variables par variable est toujours respectée. 2 variables sont liées au décès uniquement : "TEMP1" "LIVERHX" 34 variables sont liées au décès et au traitement par cathétérisme : "RESP" "CARD" "NEURO" "GASTR" "HEMA" "SEPS" "CAT1" "CAT2" "CA"  
"DAS2D3PC" "DNR1" "SURV2MD1" "APS1" "SCOMA1" "WTKILO1" "MEANBP1" "PAFI1" "PACO21"  
"PH1" "WBLC1" "HEMA1" "CREA1" "BILI1" "ALB1" "CARDIOHX" "CHFHX" "DEMENTHX" "PSYCHHX" "CHRPULHX" "GIBLEDHX" "MALIGHX" "AGE" "INCOME" "NINSCLAS" 10 variables sontliées uniquement au cathétérisme : "RENAL" "TRAUMA" "RESP1" "HRT1" "SOD1" "IMMUNHX" "TRANSHX" "AMIHX" "SEX" "EDU"

# Score de Propension

Afin d'équilibrer les groupes, je vais calculer un score de propension qui est la propabilité pour chaque patient d'être traité par la sonde de swan ganz. Pour cela je crée un modèle linéaire avec comme outcome le traitement par swan ganz et comme variables explicatives toutes les variables liées au décès à 30j uniquement(facteurs pronostiques) ou au décès et au traitement(facteurs de confusion). Je n'intègre pas les variables liées uniquement au traitement pour ne pas perdre de puissance à la fin.

Condition de validité : C'est une régression logistique, je dois avoir 5 à 10 évènement par variable. J'ai 36 variables explicatives dans le score de propension dont 5 variables qualitatives : 9 classes pour CAT1, 7 classes pour CAT2, 3 classes pour CA, 4 classes pour INCOME, 6 classes pour NINSCLAS. Ces variables qualitatives seront donc transformées en 9+7+3+4+6-5= 24 variables binaires. Soit l'équivalent de 36-5+24 = 55 variables dans le modèle logistique pour 2184 évènements. J'ai donc plus de 10 évènements par variable explicative, les conditions de validité sont respectées.

`stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.

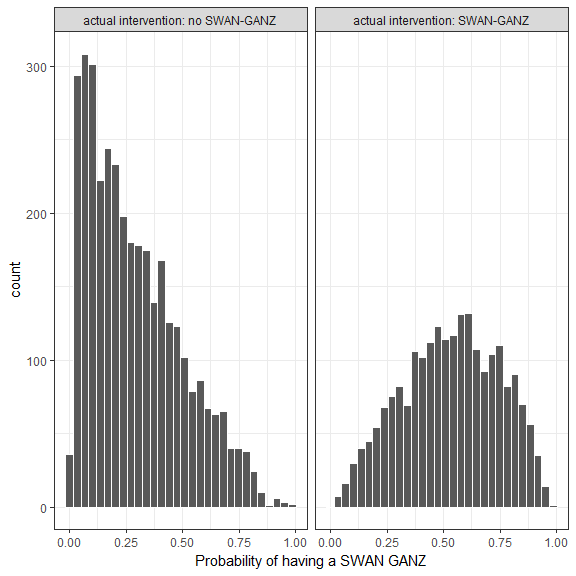


Figure x. Distribution de la probablité d'être cathétérisé selon qu'ils ont effectivement été cathétérisé ou non.

On voit que la distribution du score n'est pas la même parmi les patients ayant été cathétérisé et ceux n'ayant pas été cathétérisé : beaucoup de patient n'ayant pas été cathétérisé avait une faible probabilité d'être cathétérisé (distribution en L), alors que la distribution est plutôt en cloche centré autour de 0.5(distribution en cloche). On comprend donc que nécessairement l'appariemment va supprimer de nombreux patients.

## nombre de patients gardés

# Appariemment sur le score de propension

A partir du score de propension, je vais apparier un sujet cathétériser avec un sujet non cathétérisé ayant un score de propension proche c'est à dire une probabilité d'être cathétérisé proche. Le sujets non appariés sont écartés de l'analyse. Les sujets devrait donc se ressembler dans les 2 groupes. J'utilise un caliper de 0.2, c'est à dire un seuil d'appariemme,t de 0.2 x sd(logit(score de propension)).

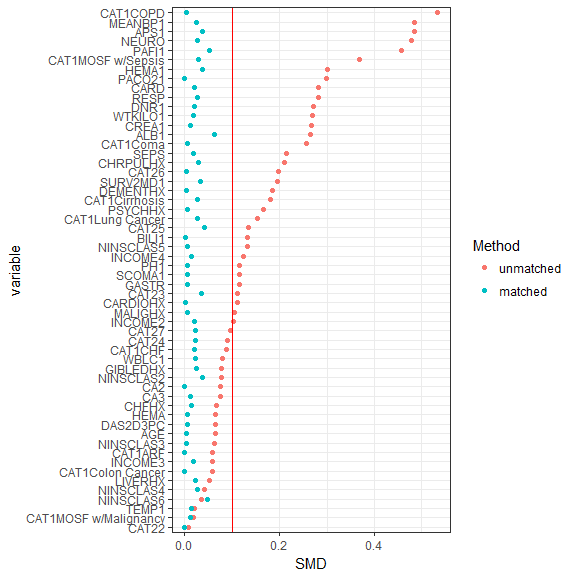
# Vérification de l'équilibre des variables entre les deux groupes après appariemment

Je transforme les qualitatives en binaires.

\*\*\*\*\* (V1) CARDIOHX \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.20421 0.2065   
mean control.......... 0.15967 0.20841   
std mean diff......... 11.046 -0.4722   
  
mean raw eQQ diff..... 0.044414 0.001912   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.02227 0.00095602   
med eCDF diff........ 0.02227 0.00095602   
max eCDF diff........ 0.044539 0.001912   
  
var ratio (Tr/Co)..... 1.2114 0.99322   
T-test p-value........ 2.6753e-05 0.89485   
  
  
\*\*\*\*\* (V2) CHFHX \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.1946 0.19949   
mean control.......... 0.16784 0.21287   
std mean diff......... 6.7572 -3.3482   
  
mean raw eQQ diff..... 0.027015 0.013384   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.013379 0.0066922   
med eCDF diff........ 0.013379 0.0066922   
max eCDF diff........ 0.026757 0.013384   
  
var ratio (Tr/Co)..... 1.1223 0.95306   
T-test p-value........ 0.01118 0.35196   
  
  
\*\*\*\*\* (V3) DEMENTHX \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.069139 0.072658   
mean control.......... 0.11631 0.069471   
std mean diff......... -18.588 1.2273   
  
mean raw eQQ diff..... 0.047161 0.0031867   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.023583 0.0015934   
med eCDF diff........ 0.023583 0.0015934   
max eCDF diff........ 0.047166 0.0031867   
  
var ratio (Tr/Co)..... 0.6263 1.0423   
T-test p-value........ 7.3188e-10 0.72949   
  
  
\*\*\*\*\* (V4) PSYCHHX \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.045788 0.051625   
mean control.......... 0.080541 0.052263   
std mean diff......... -16.623 -0.28795   
  
mean raw eQQ diff..... 0.034799 0.00063735   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.017377 0.00031867   
med eCDF diff........ 0.017377 0.00031867   
max eCDF diff........ 0.034753 0.00063735   
  
var ratio (Tr/Co)..... 0.59009 0.98847   
T-test p-value........ 5.6902e-08 0.93558   
  
  
\*\*\*\*\* (V5) CHRPULHX \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.14423 0.1536   
mean control.......... 0.21797 0.15424   
std mean diff......... -20.983 -0.17671   
  
mean raw eQQ diff..... 0.073718 0.00063735   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.036868 0.00031867   
med eCDF diff........ 0.036868 0.00031867   
max eCDF diff........ 0.073736 0.00063735   
  
var ratio (Tr/Co)..... 0.72423 0.99662   
T-test p-value........ 6.3505e-13 0.96008   
  
  
\*\*\*\*\* (V6) LIVERHX \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.062271 0.066284   
mean control.......... 0.074627 0.063098   
std mean diff......... -5.112 1.2805   
  
mean raw eQQ diff..... 0.012363 0.0031867   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.0061779 0.0015934   
med eCDF diff........ 0.0061779 0.0015934   
max eCDF diff........ 0.012356 0.0031867   
  
var ratio (Tr/Co)..... 0.84572 1.0469   
T-test p-value........ 0.069159 0.71467   
  
  
\*\*\*\*\* (V7) GIBLEDHX \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.024725 0.028681   
mean control.......... 0.036891 0.028043   
std mean diff......... -7.8326 0.38174   
  
mean raw eQQ diff..... 0.011905 0.00063735   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.0060829 0.00031867   
med eCDF diff........ 0.0060829 0.00031867   
max eCDF diff........ 0.012166 0.00063735   
  
var ratio (Tr/Co)..... 0.67881 1.0221   
T-test p-value........ 0.0080415 0.9156   
  
  
\*\*\*\*\* (V8) MALIGHX \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.2033 0.22116   
mean control.......... 0.24556 0.22881   
std mean diff......... -10.5 -1.8422   
  
mean raw eQQ diff..... 0.042125 0.0076482   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.021134 0.0038241   
med eCDF diff........ 0.021134 0.0038241   
max eCDF diff........ 0.042268 0.0076482   
  
var ratio (Tr/Co)..... 0.87441 0.97616   
T-test p-value........ 0.00017202 0.6036   
  
  
\*\*\*\*\* (V9) AGE \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 60.75 60.606   
mean control.......... 61.761 60.723   
std mean diff......... -6.4686 -0.74283   
  
mean raw eQQ diff..... 1.684 1.2541   
med raw eQQ diff..... 1.582 0.88098   
max raw eQQ diff..... 4.797 4.383   
  
mean eCDF diff........ 0.028456 0.017631   
med eCDF diff........ 0.022275 0.015296   
max eCDF diff........ 0.070272 0.051625   
  
var ratio (Tr/Co)..... 0.81749 0.83704   
T-test p-value........ 0.022435 0.83881   
KS Bootstrap p-value.. < 2.22e-16 0.024   
KS Naive p-value...... 3.1678e-06 0.030547   
KS Statistic.......... 0.070272 0.051625   
  
  
\*\*\*\*\* (V10) SURV2MD1 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.56846 0.58945   
mean control.......... 0.60721 0.58812   
std mean diff......... -19.539 0.68543   
  
mean raw eQQ diff..... 0.038596 0.0091617   
med raw eQQ diff..... 0.039978 0.0069551   
max raw eQQ diff..... 0.064 0.033999   
  
mean eCDF diff........ 0.047529 0.01098   
med eCDF diff........ 0.041643 0.010198   
max eCDF diff........ 0.095698 0.035054   
  
var ratio (Tr/Co)..... 1.0663 0.91482   
T-test p-value........ 4.1656e-13 0.84913   
KS Bootstrap p-value.. < 2.22e-16 0.3   
KS Naive p-value...... 3.5003e-11 0.29001   
KS Statistic.......... 0.095698 0.035054   
  
  
\*\*\*\*\* (V11) DAS2D3PC \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 20.701 20.584   
mean control.......... 20.371 20.524   
std mean diff......... 6.5434 1.1894   
  
mean raw eQQ diff..... 0.49681 0.36021   
med raw eQQ diff..... 0.28516 0.12891   
max raw eQQ diff..... 3.8516 3.8516   
  
mean eCDF diff........ 0.024966 0.011329   
med eCDF diff........ 0.013749 0.0070108   
max eCDF diff........ 0.066851 0.047801   
  
var ratio (Tr/Co)..... 0.8429 0.85102   
T-test p-value........ 0.020104 0.74884   
KS Bootstrap p-value.. < 2.22e-16 0.046   
KS Naive p-value...... 1.1267e-05 0.055467   
KS Statistic.......... 0.066851 0.047801   
  
  
\*\*\*\*\* (V12) APS1 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 60.739 56.892   
mean control.......... 50.934 56.967   
std mean diff......... 48.371 -0.39331   
  
mean raw eQQ diff..... 9.8324 0.7833   
med raw eQQ diff..... 10 1   
max raw eQQ diff..... 14 18   
  
mean eCDF diff........ 0.079748 0.0066007   
med eCDF diff........ 0.057146 0.0057361   
max eCDF diff........ 0.21274 0.026131   
  
var ratio (Tr/Co)..... 1.1609 0.97714   
T-test p-value........ < 2.22e-16 0.90707   
KS Bootstrap p-value.. < 2.22e-16 0.524   
KS Naive p-value...... < 2.22e-16 0.65767   
KS Statistic.......... 0.21274 0.026131   
  
  
\*\*\*\*\* (V13) SCOMA1 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 18.973 18.574   
mean control.......... 22.253 18.697   
std mean diff......... -11.604 -0.43915   
  
mean raw eQQ diff..... 3.266 0.92862   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 34 28   
  
mean eCDF diff........ 0.03459 0.0086332   
med eCDF diff........ 0.027772 0.0057361   
max eCDF diff........ 0.068259 0.020395   
  
var ratio (Tr/Co)..... 0.81163 0.95211   
T-test p-value........ 4.3792e-05 0.90072   
KS Bootstrap p-value.. < 2.22e-16 0.498   
KS Naive p-value...... 6.7334e-06 0.89992   
KS Statistic.......... 0.068259 0.020395   
  
  
\*\*\*\*\* (V14) MEANBP1 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 67.906 73.192   
mean control.......... 84.471 72.927   
std mean diff......... -48.388 0.74607   
  
mean raw eQQ diff..... 16.558 2.4254   
med raw eQQ diff..... 11 1   
max raw eQQ diff..... 49 37   
  
mean eCDF diff........ 0.092063 0.013598   
med eCDF diff........ 0.069097 0.01211   
max eCDF diff........ 0.21112 0.036966   
  
var ratio (Tr/Co)..... 0.76118 1.0077   
T-test p-value........ < 2.22e-16 0.82733   
KS Bootstrap p-value.. < 2.22e-16 0.194   
KS Naive p-value...... < 2.22e-16 0.23398   
KS Statistic.......... 0.21112 0.036966   
  
  
\*\*\*\*\* (V15) WBLC1 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 16.266 15.919   
mean control.......... 15.263 15.922   
std mean diff......... 7.9869 -0.025438   
  
mean raw eQQ diff..... 1.0778 0.41576   
med raw eQQ diff..... 0.89844 0.10156   
max raw eQQ diff..... 71 92   
  
mean eCDF diff........ 0.019048 0.0047231   
med eCDF diff........ 0.012898 0.0038241   
max eCDF diff........ 0.056097 0.020395   
  
var ratio (Tr/Co)..... 1.2088 1.0854   
T-test p-value........ 0.0023905 0.99422   
KS Bootstrap p-value.. < 2.22e-16 0.868   
KS Naive p-value...... 0.00040252 0.89992   
KS Statistic.......... 0.056097 0.020395   
  
  
\*\*\*\*\* (V16) TEMP1 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 37.595 37.636   
mean control.......... 37.633 37.578   
std mean diff......... -2.0869 3.326   
  
mean raw eQQ diff..... 0.07974 0.1371   
med raw eQQ diff..... 0 0.10156   
max raw eQQ diff..... 1.5977 3.4023   
  
mean eCDF diff........ 0.0069645 0.012263   
med eCDF diff........ 0.0037106 0.0082855   
max eCDF diff........ 0.028031 0.035692   
  
var ratio (Tr/Co)..... 1.1024 0.81985   
T-test p-value........ 0.43463 0.37487   
KS Bootstrap p-value.. 0.164 0.224   
KS Naive p-value...... 0.23842 0.27038   
KS Statistic.......... 0.028031 0.035692   
  
  
\*\*\*\*\* (V17) PAFI1 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 192.43 208.69   
mean control.......... 240.63 210.88   
std mean diff......... -45.664 -2.015   
  
mean raw eQQ diff..... 48.13 3.6892   
med raw eQQ diff..... 46.312 2.5   
max raw eQQ diff..... 241.88 94.25   
  
mean eCDF diff........ 0.11148 0.0095025   
med eCDF diff........ 0.12874 0.0044614   
max eCDF diff........ 0.18158 0.036329   
  
var ratio (Tr/Co)..... 0.8184 1.0242   
T-test p-value........ < 2.22e-16 0.55171   
KS Bootstrap p-value.. < 2.22e-16 0.254   
KS Naive p-value...... < 2.22e-16 0.25168   
KS Statistic.......... 0.18158 0.036329   
  
  
\*\*\*\*\* (V18) ALB1 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 2.966 3.0343   
mean control.......... 3.1613 3.0252   
std mean diff......... -26.445 1.2971   
  
mean raw eQQ diff..... 0.19482 0.056082   
med raw eQQ diff..... 0.2998 0   
max raw eQQ diff..... 1.0996 0.30029   
  
mean eCDF diff........ 0.037204 0.011166   
med eCDF diff........ 0.0235 0.0035054   
max eCDF diff........ 0.09896 0.06246   
  
var ratio (Tr/Co)..... 1.2559 1.0309   
T-test p-value........ < 2.22e-16 0.71146   
KS Bootstrap p-value.. < 2.22e-16 < 2.22e-16   
KS Naive p-value...... 6.2834e-12 0.0043921   
KS Statistic.......... 0.09896 0.06246   
  
  
\*\*\*\*\* (V19) HEMA1 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 30.516 30.941   
mean control.......... 32.744 30.865   
std mean diff......... -30.101 1.0083   
  
mean raw eQQ diff..... 2.3089 0.62188   
med raw eQQ diff..... 2 0.5   
max raw eQQ diff..... 6.6016 6.2969   
  
mean eCDF diff........ 0.051603 0.015393   
med eCDF diff........ 0.025675 0.012747   
max eCDF diff........ 0.14817 0.039516   
  
var ratio (Tr/Co)..... 0.7201 0.93165   
T-test p-value........ < 2.22e-16 0.7822   
KS Bootstrap p-value.. < 2.22e-16 0.122   
KS Naive p-value...... < 2.22e-16 0.17248   
KS Statistic.......... 0.14817 0.039516   
  
  
\*\*\*\*\* (V20) BILI1 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 2.7057 2.4129   
mean control.......... 1.9973 2.4612   
std mean diff......... 13.291 -0.99086   
  
mean raw eQQ diff..... 0.73006 0.17042   
med raw eQQ diff..... 0.10986 0   
max raw eQQ diff..... 9.4004 11.199   
  
mean eCDF diff........ 0.024365 0.0045228   
med eCDF diff........ 0.019348 0.0025494   
max eCDF diff........ 0.11186 0.085405   
  
var ratio (Tr/Co)..... 1.4504 0.81885   
T-test p-value........ 2.0365e-07 0.7936   
KS Bootstrap p-value.. < 2.22e-16 < 2.22e-16   
KS Naive p-value...... 3.9968e-15 2.1422e-05   
KS Statistic.......... 0.11186 0.085405   
  
  
\*\*\*\*\* (V21) CREA1 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 2.4734 2.2715   
mean control.......... 1.9236 2.2821   
std mean diff......... 26.781 -0.55411   
  
mean raw eQQ diff..... 0.58035 0.19011   
med raw eQQ diff..... 0.40015 0.1001   
max raw eQQ diff..... 8.0977 4.8984   
  
mean eCDF diff........ 0.039684 0.013193   
med eCDF diff........ 0.0037401 0.0098789   
max eCDF diff........ 0.20108 0.044614   
  
var ratio (Tr/Co)..... 1.0276 0.71618   
T-test p-value........ < 2.22e-16 0.88263   
KS Bootstrap p-value.. < 2.22e-16 0.036   
KS Naive p-value...... < 2.22e-16 0.088042   
KS Statistic.......... 0.20108 0.044614   
  
  
\*\*\*\*\* (V22) PACO21 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 36.661 37.295   
mean control.......... 39.721 37.554   
std mean diff......... -29.879 -2.4357   
  
mean raw eQQ diff..... 3.0837 0.49211   
med raw eQQ diff..... 2 0   
max raw eQQ diff..... 17 7   
  
mean eCDF diff........ 0.047051 0.008867   
med eCDF diff........ 0.043809 0.0057361   
max eCDF diff........ 0.10809 0.027406   
  
var ratio (Tr/Co)..... 0.58943 0.99687   
T-test p-value........ < 2.22e-16 0.48648   
KS Bootstrap p-value.. < 2.22e-16 0.436   
KS Naive p-value...... 3.7859e-14 0.59761   
KS Statistic.......... 0.10809 0.027406   
  
  
\*\*\*\*\* (V23) PH1 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 7.3802 7.3895   
mean control.......... 7.3935 7.3894   
std mean diff......... -11.625 0.15557   
  
mean raw eQQ diff..... 0.013163 0.0056938   
med raw eQQ diff..... 0.0097656 0   
max raw eQQ diff..... 0.13086 0.13086   
  
mean eCDF diff........ 0.014537 0.0068165   
med eCDF diff........ 0.0043743 0.0044614   
max eCDF diff........ 0.087984 0.028681   
  
var ratio (Tr/Co)..... 1.1322 0.98916   
T-test p-value........ 1.2524e-05 0.96546   
KS Bootstrap p-value.. < 2.22e-16 0.366   
KS Naive p-value...... 1.6156e-09 0.53876   
KS Statistic.......... 0.087984 0.028681   
  
  
\*\*\*\*\* (V24) WTKILO1 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 77.802 75.806   
mean control.......... 72.374 75.94   
std mean diff......... 26.936 -0.71713   
  
mean raw eQQ diff..... 5.499 1.891   
med raw eQQ diff..... 5.6 1.3   
max raw eQQ diff..... 39.5 43.6   
  
mean eCDF diff........ 0.057958 0.016404   
med eCDF diff........ 0.04963 0.014659   
max eCDF diff........ 0.13201 0.045252   
  
var ratio (Tr/Co)..... 0.96282 0.7616   
T-test p-value........ < 2.22e-16 0.8477   
KS Bootstrap p-value.. < 2.22e-16 0.078   
KS Naive p-value...... < 2.22e-16 0.080476   
KS Statistic.......... 0.13201 0.045252   
  
  
\*\*\*\*\* (V25) DNR1 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.070971 0.083493   
mean control.......... 0.14052 0.088591   
std mean diff......... -27.081 -1.8426   
  
mean raw eQQ diff..... 0.069597 0.0050988   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.034777 0.0025494   
med eCDF diff........ 0.034777 0.0025494   
max eCDF diff........ 0.069553 0.0050988   
  
var ratio (Tr/Co)..... 0.54601 0.94772   
T-test p-value........ < 2.22e-16 0.61434   
  
  
\*\*\*\*\* (V26) RESP \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.28938 0.33333   
mean control.......... 0.41707 0.32823   
std mean diff......... -28.151 1.0813   
  
mean raw eQQ diff..... 0.12775 0.0050988   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.063844 0.0025494   
med eCDF diff........ 0.063844 0.0025494   
max eCDF diff........ 0.12769 0.0050988   
  
var ratio (Tr/Co)..... 0.84597 1.0078   
T-test p-value........ < 2.22e-16 0.75408   
  
  
\*\*\*\*\* (V27) CARD \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.42308 0.38942   
mean control.......... 0.28358 0.40153   
std mean diff......... 28.229 -2.4826   
  
mean raw eQQ diff..... 0.13965 0.01211   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.069747 0.0060548   
med eCDF diff........ 0.069747 0.0060548   
max eCDF diff........ 0.13949 0.01211   
  
var ratio (Tr/Co)..... 1.2016 0.98947   
T-test p-value........ < 2.22e-16 0.47491   
  
  
\*\*\*\*\* (V28) NEURO \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.054029 0.069471   
mean control.......... 0.16193 0.059273   
std mean diff......... -47.715 4.0095   
  
mean raw eQQ diff..... 0.1076 0.010198   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.053948 0.0050988   
med eCDF diff........ 0.053948 0.0050988   
max eCDF diff........ 0.1079 0.010198   
  
var ratio (Tr/Co)..... 0.37669 1.1593   
T-test p-value........ < 2.22e-16 0.21973   
  
  
\*\*\*\*\* (V29) GASTR \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.19231 0.17846   
mean control.......... 0.147 0.17463   
std mean diff......... 11.493 0.99841   
  
mean raw eQQ diff..... 0.04533 0.0038241   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.022653 0.001912   
med eCDF diff........ 0.022653 0.001912   
max eCDF diff........ 0.045307 0.0038241   
  
var ratio (Tr/Co)..... 1.2389 1.0172   
T-test p-value........ 1.1569e-05 0.77829   
  
  
\*\*\*\*\* (V30) HEMA \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.052656 0.060548   
mean control.......... 0.067305 0.063098   
std mean diff......... -6.5575 -1.0686   
  
mean raw eQQ diff..... 0.014652 0.0025494   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.0073247 0.0012747   
med eCDF diff........ 0.0073247 0.0012747   
max eCDF diff........ 0.014649 0.0025494   
  
var ratio (Tr/Co)..... 0.79477 0.96221   
T-test p-value........ 0.021435 0.7604   
  
  
\*\*\*\*\* (V31) SEPS \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.23626 0.20013   
mean control.......... 0.14503 0.21224   
std mean diff......... 21.473 -3.0257   
  
mean raw eQQ diff..... 0.091117 0.01211   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.045617 0.0060548   
med eCDF diff........ 0.045617 0.0060548   
max eCDF diff........ 0.091234 0.01211   
  
var ratio (Tr/Co)..... 1.4555 0.95744   
T-test p-value........ < 2.22e-16 0.39024   
  
  
\*\*\*\*\* (V32) CAT1\_2 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.095696 0.10962   
mean control.......... 0.069558 0.11727   
std mean diff......... 8.8832 -2.4473   
  
mean raw eQQ diff..... 0.026099 0.0076482   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.013069 0.0038241   
med eCDF diff........ 0.013069 0.0038241   
max eCDF diff........ 0.026138 0.0076482   
  
var ratio (Tr/Co)..... 1.3374 0.94288   
T-test p-value........ 0.00059649 0.49694   
  
  
\*\*\*\*\* (V33) CAT1\_3 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.022436 0.028681   
mean control.......... 0.049282 0.028043   
std mean diff......... -18.123 0.38174   
  
mean raw eQQ diff..... 0.026557 0.00063735   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.013423 0.00031867   
med eCDF diff........ 0.013423 0.00031867   
max eCDF diff........ 0.026846 0.00063735   
  
var ratio (Tr/Co)..... 0.46819 1.0221   
T-test p-value........ 2.6922e-08 0.91154   
  
  
\*\*\*\*\* (V34) CAT1\_4 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.00045788 0.00063735   
mean control.......... 0.0016897 0   
std mean diff......... -5.7566 2.5246   
  
mean raw eQQ diff..... 0.0013736 0.00063735   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.00061589 0.00031867   
med eCDF diff........ 0.00061589 0.00031867   
max eCDF diff........ 0.0012318 0.00063735   
  
var ratio (Tr/Co)..... 0.27137 Inf   
T-test p-value........ 0.13667 0.31731   
  
  
\*\*\*\*\* (V35) CAT1\_5 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.043498 0.049076   
mean control.......... 0.096029 0.054175   
std mean diff......... -25.748 -2.3595   
  
mean raw eQQ diff..... 0.052656 0.0050988   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.026266 0.0025494   
med eCDF diff........ 0.026266 0.0025494   
max eCDF diff........ 0.052531 0.0050988   
  
var ratio (Tr/Co)..... 0.47937 0.91077   
T-test p-value........ 1.9984e-15 0.50204   
  
  
\*\*\*\*\* (V36) CAT1\_6 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.026557 0.036329   
mean control.......... 0.11236 0.036329   
std mean diff......... -53.355 0   
  
mean raw eQQ diff..... 0.085623 0   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 0   
  
mean eCDF diff........ 0.042903 0   
med eCDF diff........ 0.042903 0   
max eCDF diff........ 0.085806 0   
  
var ratio (Tr/Co)..... 0.25924 1   
T-test p-value........ < 2.22e-16 1   
  
  
\*\*\*\*\* (V37) CAT1\_7 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.0022894 0.0031867   
mean control.......... 0.0095748 0.00063735   
std mean diff......... -15.24 4.5219   
  
mean raw eQQ diff..... 0.007326 0.0025494   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.0036427 0.0012747   
med eCDF diff........ 0.0036427 0.0012747   
max eCDF diff........ 0.0072854 0.0025494   
  
var ratio (Tr/Co)..... 0.24091 4.9872   
T-test p-value........ 0.00015944 0.10238   
  
  
\*\*\*\*\* (V38) CAT1\_8 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.072344 0.078394   
mean control.......... 0.067868 0.083493   
std mean diff......... 1.7275 -1.8963   
  
mean raw eQQ diff..... 0.0045788 0.0050988   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.0022381 0.0025494   
med eCDF diff........ 0.0022381 0.0025494   
max eCDF diff........ 0.0044761 0.0050988   
  
var ratio (Tr/Co)..... 1.061 0.94415   
T-test p-value........ 0.5207 0.60258   
  
  
\*\*\*\*\* (V39) CAT1\_9 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.32051 0.25621   
mean control.......... 0.14841 0.25048   
std mean diff......... 36.87 1.3136   
  
mean raw eQQ diff..... 0.17216 0.0057361   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.086052 0.0028681   
med eCDF diff........ 0.086052 0.0028681   
max eCDF diff........ 0.1721 0.0057361   
  
var ratio (Tr/Co)..... 1.7235 1.0151   
T-test p-value........ < 2.22e-16 0.6967   
  
  
\*\*\*\*\* (V40) CAT2\_2 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.00045788 0.00063735   
mean control.......... 0.00028161 0.00063735   
std mean diff......... 0.82374 0   
  
mean raw eQQ diff..... 0 0   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 0 0   
  
mean eCDF diff........ 8.8132e-05 0   
med eCDF diff........ 8.8132e-05 0   
max eCDF diff........ 0.00017626 0   
  
var ratio (Tr/Co)..... 1.6259 1   
T-test p-value........ 0.743 1   
  
  
\*\*\*\*\* (V41) CAT2\_3 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.0091575 0.0095602   
mean control.......... 0.019713 0.0089229   
std mean diff......... -11.078 0.65477   
  
mean raw eQQ diff..... 0.010531 0.00063735   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.0052776 0.00031867   
med eCDF diff........ 0.0052776 0.00031867   
max eCDF diff........ 0.010555 0.00063735   
  
var ratio (Tr/Co)..... 0.46963 1.0707   
T-test p-value........ 0.0006621 0.85271   
  
  
\*\*\*\*\* (V42) CAT2\_4 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.00091575 0.0012747   
mean control.......... 0.0036609 0.00063735   
std mean diff......... -9.0737 1.7857   
  
mean raw eQQ diff..... 0.0027473 0.00063735   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.0013726 0.00031867   
med eCDF diff........ 0.0013726 0.00031867   
max eCDF diff........ 0.0027452 0.00063735   
  
var ratio (Tr/Co)..... 0.25087 1.9987   
T-test p-value........ 0.022501 0.56374   
  
  
\*\*\*\*\* (V43) CAT2\_5 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.026557 0.034417   
mean control.......... 0.048155 0.034417   
std mean diff......... -13.43 0   
  
mean raw eQQ diff..... 0.02152 0   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 0   
  
mean eCDF diff........ 0.010799 0   
med eCDF diff........ 0.010799 0   
max eCDF diff........ 0.021599 0   
  
var ratio (Tr/Co)..... 0.56409 1   
T-test p-value........ 1.4427e-05 1   
  
  
\*\*\*\*\* (V44) CAT2\_6 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.19231 0.16571   
mean control.......... 0.11433 0.16571   
std mean diff......... 19.78 0   
  
mean raw eQQ diff..... 0.077839 0   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 0   
  
mean eCDF diff........ 0.038987 0   
med eCDF diff........ 0.038987 0   
max eCDF diff........ 0.077974 0   
  
var ratio (Tr/Co)..... 1.5342 1   
T-test p-value........ 7.3275e-15 1   
  
  
\*\*\*\*\* (V45) CAT2\_7 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.76557 0.7833   
mean control.......... 0.80625 0.78394   
std mean diff......... -9.6012 -0.15465   
  
mean raw eQQ diff..... 0.040293 0.00063735   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.020342 0.00031867   
med eCDF diff........ 0.020342 0.00031867   
max eCDF diff........ 0.040684 0.00063735   
  
var ratio (Tr/Co)..... 1.1491 1.0021   
T-test p-value........ 0.00029645 0.96526   
  
  
\*\*\*\*\* (V46) CA\_2 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.056319 0.059911   
mean control.......... 0.0735 0.068196   
std mean diff......... -7.4512 -3.4902   
  
mean raw eQQ diff..... 0.016941 0.0082855   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.0085909 0.0041428   
med eCDF diff........ 0.0085909 0.0041428   
max eCDF diff........ 0.017182 0.0082855   
  
var ratio (Tr/Co)..... 0.78058 0.88632   
T-test p-value........ 0.0092348 0.34436   
  
  
\*\*\*\*\* (V47) CA\_3 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.15293 0.1689   
mean control.......... 0.17967 0.16699   
std mean diff......... -7.427 0.51018   
  
mean raw eQQ diff..... 0.026557 0.001912   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.013369 0.00095602   
med eCDF diff........ 0.013369 0.00095602   
max eCDF diff........ 0.026737 0.001912   
  
var ratio (Tr/Co)..... 0.87908 1.0091   
T-test p-value........ 0.0077861 0.88616   
  
  
\*\*\*\*\* (V48) INCOME\_2 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.17995 0.16125   
mean control.......... 0.14081 0.16125   
std mean diff......... 10.187 0   
  
mean raw eQQ diff..... 0.039377 0   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 0   
  
mean eCDF diff........ 0.01957 0   
med eCDF diff........ 0.01957 0   
max eCDF diff........ 0.03914 0   
  
var ratio (Tr/Co)..... 1.22 1   
T-test p-value........ 0.00010535 1   
  
  
\*\*\*\*\* (V49) INCOME\_3 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.088828 0.079669   
mean control.......... 0.072374 0.077757   
std mean diff......... 5.7822 0.7059   
  
mean raw eQQ diff..... 0.016484 0.001912   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.0082269 0.00095602   
med eCDF diff........ 0.0082269 0.00095602   
max eCDF diff........ 0.016454 0.001912   
  
var ratio (Tr/Co)..... 1.2058 1.0225   
T-test p-value........ 0.027933 0.84288   
  
  
\*\*\*\*\* (V50) INCOME\_4 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.52427 0.54685   
mean control.......... 0.58603 0.55258   
std mean diff......... -12.365 -1.1519   
  
mean raw eQQ diff..... 0.061813 0.0057361   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.030882 0.0028681   
med eCDF diff........ 0.030882 0.0028681   
max eCDF diff........ 0.061765 0.0057361   
  
var ratio (Tr/Co)..... 1.0283 1.0023   
T-test p-value........ 4.9842e-06 0.74524   
  
  
\*\*\*\*\* (V51) NINSCLAS\_2 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.23397 0.23454   
mean control.......... 0.26669 0.24092   
std mean diff......... -7.7248 -1.5037   
  
mean raw eQQ diff..... 0.032509 0.0063735   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.016356 0.0031867   
med eCDF diff........ 0.016356 0.0031867   
max eCDF diff........ 0.032711 0.0063735   
  
var ratio (Tr/Co)..... 0.91664 0.98172   
T-test p-value........ 0.0052471 0.67696   
  
  
\*\*\*\*\* (V52) NINSCLAS\_3 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.056319 0.064372   
mean control.......... 0.070684 0.063098   
std mean diff......... -6.23 0.51924   
  
mean raw eQQ diff..... 0.014194 0.0012747   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.0071828 0.00063735   
med eCDF diff........ 0.0071828 0.00063735   
max eCDF diff........ 0.014366 0.0012747   
  
var ratio (Tr/Co)..... 0.80922 1.0188   
T-test p-value........ 0.028239 0.8828   
  
  
\*\*\*\*\* (V53) NINSCLAS\_4 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.062271 0.059273   
mean control.......... 0.05238 0.057999   
std mean diff......... 4.0924 0.53964   
  
mean raw eQQ diff..... 0.010073 0.0012747   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.0049457 0.00063735   
med eCDF diff........ 0.0049457 0.00063735   
max eCDF diff........ 0.0098915 0.0012747   
  
var ratio (Tr/Co)..... 1.1766 1.0206   
T-test p-value........ 0.12124 0.87881   
  
  
\*\*\*\*\* (V54) NINSCLAS\_5 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.33471 0.31358   
mean control.......... 0.27232 0.30848   
std mean diff......... 13.218 1.0987   
  
mean raw eQQ diff..... 0.062271 0.0050988   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.031195 0.0025494   
med eCDF diff........ 0.031195 0.0025494   
max eCDF diff........ 0.062389 0.0050988   
  
var ratio (Tr/Co)..... 1.1239 1.009   
T-test p-value........ 7.0868e-07 0.75517   
  
  
\*\*\*\*\* (V55) NINSCLAS\_6 \*\*\*\*\*  
 Before Matching After Matching  
mean treatment........ 0.22436 0.23454   
mean control.......... 0.21008 0.22371   
std mean diff......... 3.4217 2.5563   
  
mean raw eQQ diff..... 0.014194 0.010835   
med raw eQQ diff..... 0 0   
max raw eQQ diff..... 1 1   
  
mean eCDF diff........ 0.0071387 0.0054175   
med eCDF diff........ 0.0071387 0.0054175   
max eCDF diff........ 0.014277 0.010835   
  
var ratio (Tr/Co)..... 1.0488 1.0338   
T-test p-value........ 0.20429 0.47057   
  
  
Before Matching Minimum p.value: < 2.22e-16   
Variable Name(s): AGE SURV2MD1 DAS2D3PC APS1 SCOMA1 MEANBP1 WBLC1 PAFI1 ALB1 HEMA1 BILI1 CREA1 PACO21 PH1 WTKILO1 DNR1 RESP CARD NEURO SEPS CAT1\_6 CAT1\_9 Number(s): 9 10 11 12 13 14 15 17 18 19 20 21 22 23 24 25 26 27 28 31 36 39   
  
After Matching Minimum p.value: < 2.22e-16   
Variable Name(s): ALB1 BILI1 Number(s): 18 20

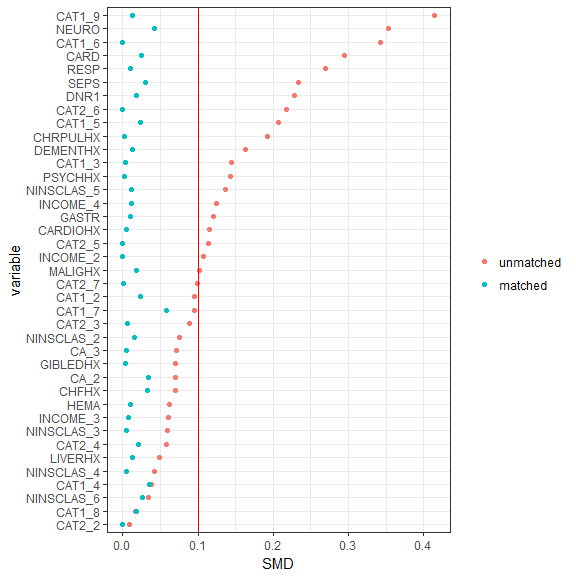
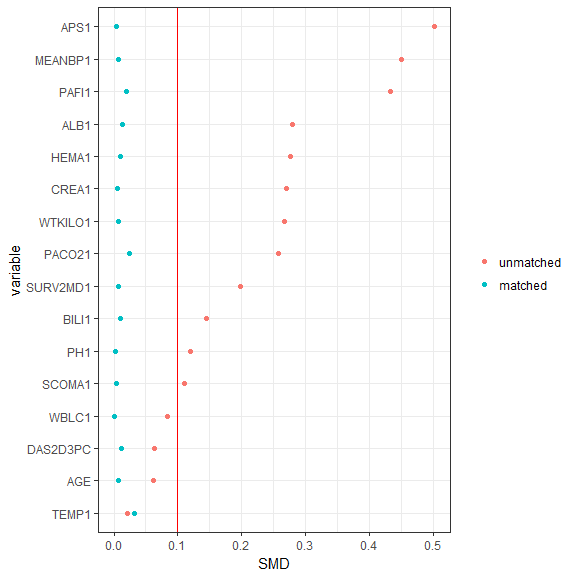
[1] "AGE"  
[1] "SURV2MD1"  
[1] "DAS2D3PC"  
[1] "APS1"  
[1] "SCOMA1"  
[1] "MEANBP1"  
[1] "WBLC1"  
[1] "TEMP1"  
[1] "PAFI1"  
[1] "ALB1"  
[1] "HEMA1"  
[1] "BILI1"  
[1] "CREA1"  
[1] "PACO21"  
[1] "PH1"  
[1] "WTKILO1"  
[1] "CARDIOHX"  
[1] "CHFHX"  
[1] "DEMENTHX"  
[1] "PSYCHHX"  
[1] "CHRPULHX"  
[1] "LIVERHX"  
[1] "GIBLEDHX"  
[1] "MALIGHX"  
[1] "DNR1"  
[1] "RESP"  
[1] "CARD"  
[1] "NEURO"  
[1] "GASTR"  
[1] "HEMA"  
[1] "SEPS"  
[1] "CAT1\_2"  
[1] "CAT1\_3"  
[1] "CAT1\_4"  
[1] "CAT1\_5"  
[1] "CAT1\_6"  
[1] "CAT1\_7"  
[1] "CAT1\_8"  
[1] "CAT1\_9"  
[1] "CAT2\_2"  
[1] "CAT2\_3"  
[1] "CAT2\_4"  
[1] "CAT2\_5"  
[1] "CAT2\_6"  
[1] "CAT2\_7"  
[1] "CA\_2"  
[1] "CA\_3"  
[1] "INCOME\_2"  
[1] "INCOME\_3"  
[1] "INCOME\_4"  
[1] "NINSCLAS\_2"  
[1] "NINSCLAS\_3"  
[1] "NINSCLAS\_4"  
[1] "NINSCLAS\_5"  
[1] "NINSCLAS\_6"

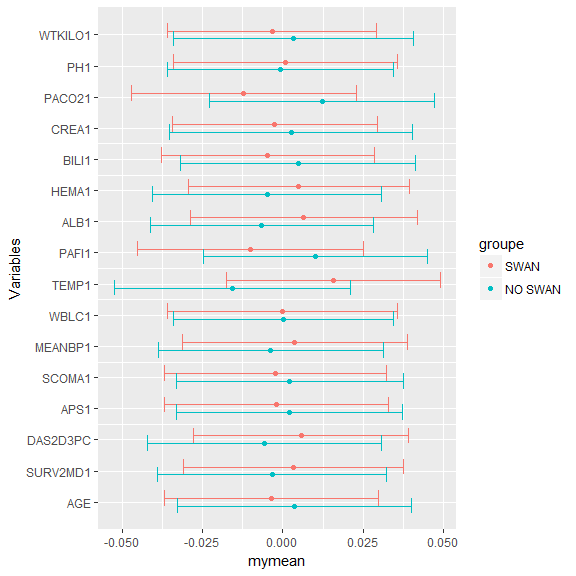
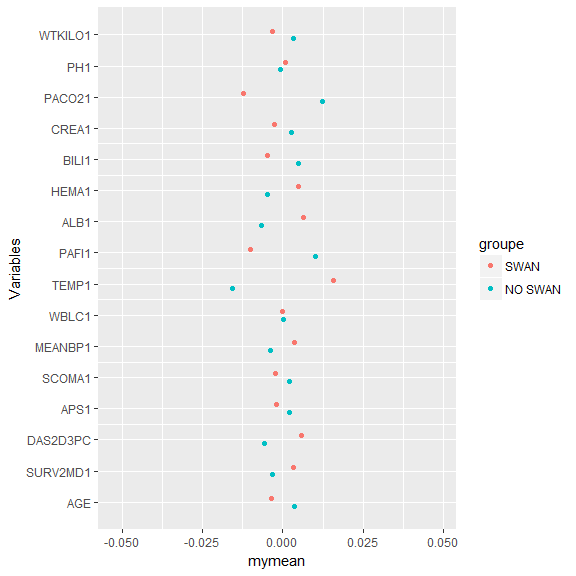
Saving 6 x 6 in image



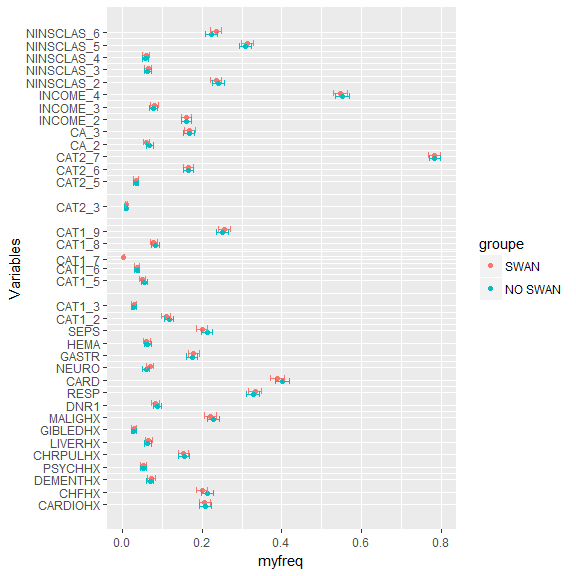
[1] -0.06135247

[1] -0.06468618





[1] 1  
[1] 2  
[1] 3  
[1] 4  
[1] 5  
[1] 6  
[1] 7  
[1] 8  
[1] 9  
[1] 10  
[1] 11  
[1] 12  
[1] 13  
[1] 14  
[1] 15  
[1] 16  
[1] 17  
[1] 18  
[1] 19  
[1] 20  
[1] 21  
[1] 22  
[1] 23  
[1] 24  
[1] 25  
[1] 26  
[1] 27  
[1] 28  
[1] 29  
[1] 30  
[1] 31  
[1] 32  
[1] 33  
[1] 34  
[1] 35  
[1] 36  
[1] 37  
[1] 38  
[1] 39



Tout d'abord, je transforme mes variables qualitatives en n-k variables binaires, k étant le nombre de classe de la variable qualitative.  
J'ai ensuite plusieurs méthodes possibles pour regarder si l'appariemment a équilibré la distribution des variables dans le groupe traité et non traité : - je peux regarder la moyenne de chaque variable en fonction du score de popension. Je vois donc pour des patients de ces 2 groupes ayant la même probabilité d'être traité, si la moyenne de la variable est semblable. Bien sûr si les courbes se superposent parfaitement, on peut dire que la distribution est semblable dans les deux groupes.Si elles sont disjointes à certaines valeures d'abscisse, nous pouvont dire que pour les individus ayant tel probabilité d'être traité, la moyenne des variables diffèrent. Ainsi dans les courbes présentées en annexe, on voit ainsi que les individus avec une forte probabilité d'être traité ont une moyenne qui semble différer entre les deux groupes pour INCOME, NINCLAS, RESP et SEPS, et que les individus avec une faible probabilité d'être traité ont une moyenne qui semble différer poyr les variables RESP, HEMA et DNR1. Or il me semble qu'on ne cherche pas à ce que les individus ayant la même probabilité d'être traité soit exactement semblable 2 à 2 (même si bien sûr c'est idéal) mais plutôt que les populations des groupes traités et non traités soit homogène, comme c'est le cas lorsque l'on randomise. Il me semble donc plus pertinent de voir si globablement les distribution sont les mêmes dans les deux groupes, et les deux méthodes présentées ci dessous répondent à cette question.  
- Je regarde la différence standardisée des moyennes(smd), c'est à dire la différence entre la moyenne dans le groupe cathétérisé et la moyenne dans le groupe non cathétérisé divisé par la variance commune. J'ai séparé les schémas en variable binaire et variable quantitative non binaire pour plus de visibilité. En instaurant un seuil de smd à 0.1 comme conseillé dans la littérature, je vois que l'appariemment établi un équilibre entre les deux groupes pour toutes les variables. - Je regarde les intervalles de confiances des différentes variables, j'ai réalisé les intervalles de confiance des moyennes pour les variables quantitatives et des fréquences de 1 pour les variables binaires (qui sont en 0/1). Les variables quantitatives sont standardisées, ce qui permet une lecture plus aisée du graphique des variables quantitatives. Je ne sais pas comment réaliser un équivalent de cette standardisation avec les variables binaires et le graphique est donc moins facillement analysable. A noter que les variables pour lesquelles on n'a pas pu calculer d'intervalle de confiance pour non respect des conditions de validité (np>=5 et n(1-p)>=5) ne sont pas présente dans le graphe. On voit cependant là aussi que les variables sont plutôt bien équilibrée après appariemment.

Dans le futur, je préférerai quand même l'analyse de l'équilibre par la méthode smd, j'ai d'ailleurs pu voir qu'elle était souvent utilisée dans les publications.

# Analyses dans la population appariée

Pour répondre à la question "le cathétérisme cardiaque droit modifie-t-il la survie à 30 jours?", je peux utiliser deux méthodes différentes. Soit j'utilise une régression logistique conditionnelle, soit j'utilise un modèle de Cox. Dans les deux cas, j'utilise mon échantillon apparié et je stratifie sur le numéro de paire.

## Régression logistique conditionnelle

La variable à expliquer est la mort à 30 jours, la variable explicative est le cathétérisme cardiaque et je stratifie sur le paire.

SWAN   
1.188797

2.5 % 97.5 %  
SWAN 1.053171 1.341888

Les conditions de validité sont respectées, j'ai plus de 5 à 10 évènements par variable explicative. Le coefficient du cathétérisme est significatif (pvalue<0.01), le cathétérisme cardiaque droit a donc un effet significatif sur le risaue de décès à 30 jours. le coeffcient vaut 0.17, l'exponentiel du coefficient me permet d'obtenir l'Odds ratio du risque de décès associé au cathétérisme. L'OR est de 1.19 avec un intervalle de confiance à 95% [1.05-1.34]. Le décès n'est pas un évènement rare, je ne peux donc pas interpréter l'OR comme un RR mais je peux dire que le risque de décès est augmenté lorsque le patient est cathétérisé.

## Modèle de Cox

[1] 1569

0 1 2   
0.4380587 0.4501916 0.1117497

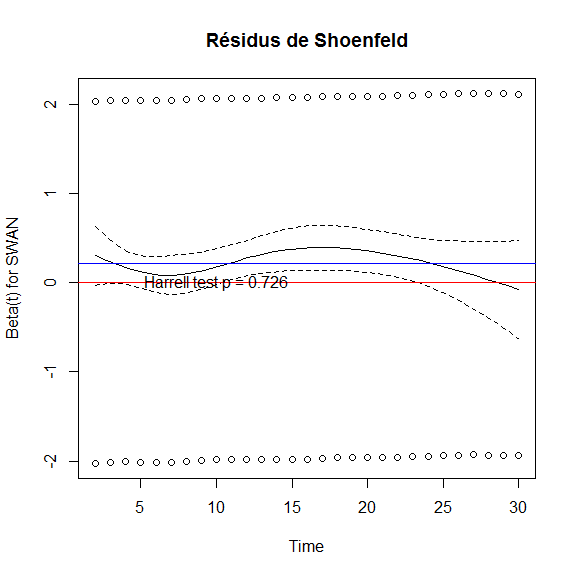
Une deuxième manière de réaliser le calcul est par modèle de cox. Dans ce cas il est problématique de prendre en compte l'appariemment par une stratification sur la paire car FE Harrell nous précise dans son livre Regression Modeling Strategies(2nd Ed) p.482 que si le nombre de strates est très grand par rapport au nombre total d'évènement, on perd en efficience(? "Loss of efficiency" dans le texte). Or on a ici 1569 paires ou strates pour 1055 évènements, ce qui peut être considéré comme un grand nombre de strates relativement au nombre d'évènements bien qu'aucun seuil ne soit précisé. Et de plus FE Harrell nous précise qu'une strate qui ne contient aucun évènement ne contribue pas à l'information et qu'une telle situation doit donc être évitée si possible. Or ici 44% des paires sont sans évènements et ne contribue pas à l'information avec donc j'imagine une perte de puissance. J'utiliserai donc l'option cluster(paire) plutôt que strata(paire) pour prendre en compte l'appariemment dans le modele de cox. Cette option prend en compte le design apparié, et on regarde alors la variance robuste et le test du score robuste. Cette méthode calcule une vraisemblance globale et non pas strate par strate comme lorsque l'on stratifie, évitant donc de perdre de l'information.

A noter que contrairement à la régression logstique conditionnelle, les sujets perdus avant 30 jours sont gérés par le modèle de cox. Cependant je préfère garder le même échantillon de patients que pour la régression logistique conditionnelle en prenant la variable décès à 30 jours modifiée telle que les patients perdus de vu avant 30 jours sont manquant pour le statut(les sujets ayant un statut manquant sont automatiquement retirés de l'analyse, comme pour le modèle de régression logistique).

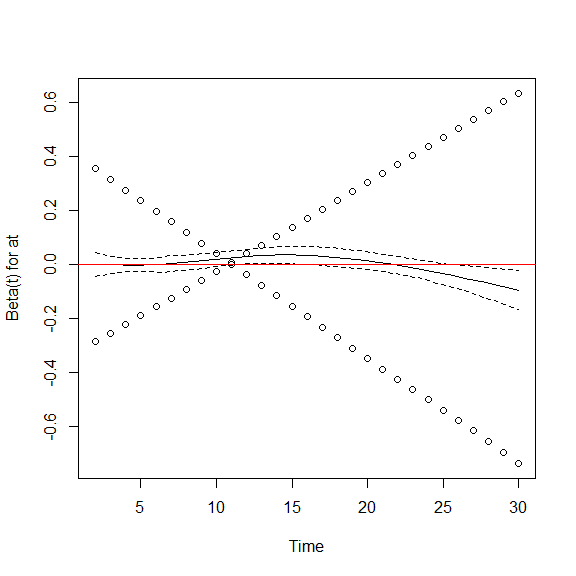
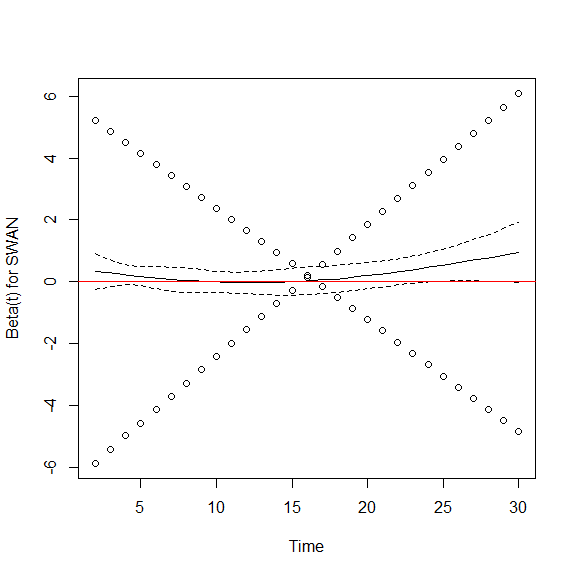
Test de Harrell

rho chisq p  
SWAN 0.0108 0.123 0.726

Test de Harrell p value: 0.726



[1] 0.7977832



pvalue   
0.0006457109

[1] "1.236" "1.233" "1.239"

L'hypothèse des risques proportionnels n'est pas vérifiée. En effet l'intervalle de confiance des résidus de shoenfeld n'incluent pas 0 en tous points. Il faut donc en toute rigueur que j'ajoute une variable SWAN dépendante du temps. Cependant je n'ai pas trouvé de transformation qui convienne, aucun coefficient de paramètre dépendant du temps n'étant significatif lorsqu'on l'ajoute au modèle (les résidus de Shoenfeld des modèles ne doivent donc même pas être regardés). Voici les transformations essayeées : log, racine carrée, \*temps, 1/temps, racine cube, carré, cube et puissance de 0.7. J'analyse donc le modèle de Cox sans ajouter de variable dépendante du temps, mais il faudra avoir en tête que le résultat est erroné.

L'ajout d'un cluster sur la paire me permet d'avoir une variance robuste prenant en compte l'appariemment. J'utilise cette variance robuste pour calculer l'intervalle de confiance à 95% du Hazard ratio: 1.236 [1.233-1.239]. Je regarde le test du score robuste pour conclure : p<0.001. Le risque de décès est donc significativement différent selon que le patient est cathétérisé ou non. Et c'est dans le sens d'un risque plus grand chez les patients cathétérisés avec un risque de décès à 30 jours multiplié par 1.236.

Je retrouve cette information graphiquement en traçant une courbe de survie en fonction du traitement par cathéterisme cardiaque droit.

## Représentation graphique de la survie à 30 jours en fonction du traitement par cathétérisme cardiaque droit

### Méthode

La méthode de Kaplan Meier permet de représenter graphiquement les courbes de survie (une courbe par groupe).

### Conditions de validité

Kaplan Meier :  
- censure indépendante de la probabilité de survenue de l'evenement  
- probabilité de survie indépendante du moment de recrutement dans l'étude  
- censure indépendante du groupe Je n'ai pas assez d'information sur le protocole de l'étude et les perdus de vu pour le vérifier.

In group CCD = 0

At 10 days, survival[95%CI] 82% [80% - 84%]  
 At 30 days, survival[95%CI] 69% [67% - 72%]

In group CCD = 1

At 10 days, survival[95%CI] 78% [76% - 81%]  
 At 30 days, survival[95%CI] 63% [61% - 66%]

Scale for 'colour' is already present. Adding another scale for  
'colour', which will replace the existing scale.

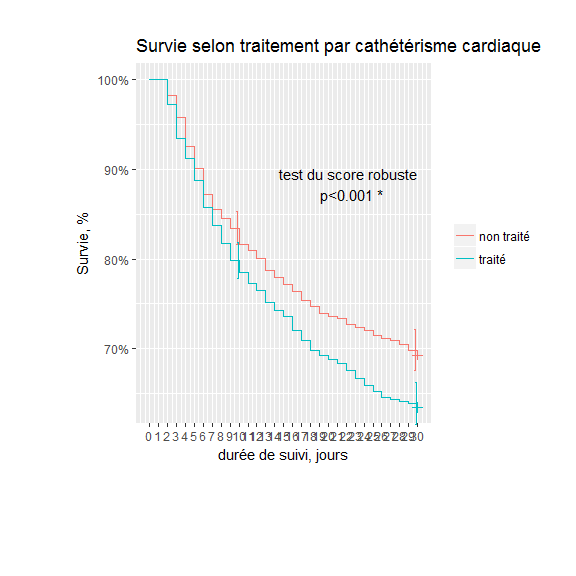


Fig X. Courbes de survie par la méthode de Kaplan Meier de la survie à 30 jours en fonction du traitement par cathétérisme cardiaque droit.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | J0 | J10 | J20 | J30 |
| n.risk | 1569 | 1308 | 1159 | 1095 |
| n.event | 0 | 288 | 414 | 482 |
| surv | 100% | 82 % | 74 % | 69 % |

Table X Table de survie des patients non traités par cathétérisme cardiaque droit

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | J0 | J10 | J20 | J30 |
| n.risk | 1569 | 1251 | 1084 | 1000 |
| n.event | 0 | 337 | 489 | 573 |
| surv | 100% | 78 % | 69 % | 63 % |

Table X Table de survie des patients traités par cathétérisme cardiaque droit

On observe 2 courbes de survie, 1 pour le groupe non traité(0) et 1 pour le groupe traité(1). A chaque temps, les "marches d'escalier" représentent les patients qui ont eu l'évènement, faisant diminuer le nombre de personnes à risque d'évènement(c'est à dire les non répondeurs). Les croix représentent à chaque temps la présence de censure (perdus de vue ou exclus vivant). Par exemple, je passe de 75% de répondeurs à J7 à un peu plus de 50% de répondeurs à J14 dans le groupe 0. Dans la table de survie (table 3) je vois que ça correspond à 16 patients ayant eu l'évènement à J14. Pour le groupe 1, je lis sur la courbe que je passe de 90% environ à 75% de répondeurs. Dans la table de survie (table 4) cela correspond à 10 patients ayant eu l'évènement. La courbe de Kaplan Meier nous permet d'observer graphiquement que la survie est meilleure lorsque les patients ne sont pas traités par cathétérisme cardiaque droit.

# Conclusion