Description de la Bases de données **Athero**

Contenu

1. Première modifications 3
   1.1. Renommage des attributs 3
   1.2. Traitement des valeurs manquantes 3
   1.3. Corrections 3
2. Objectifs 3
   2.1. Les tables de la base de données Athero 3
   2.2. Structure de la table InformationPatient 4
   2.3. Structure de la table Changes 4
3. La table InformationPatient 5
   3.1. L'attribut IDE_PATIENT_GRP 5
   3.2. L'attribut IDE_EDUC 5
   3.3. L'attribut DIAG_GEN 6
   3.4. L'attribut ALCOHOL 6
4. La table Changes 6
   4.1. Les attributs caractéristiques 6
      4.1.1. L'attribut IDE_AGE 7
      4.1.2. L'attribut IDE_DURATION 7
      4.1.3. L'attribut DIAG 7
   4.2. Les attributs de changements de comportement 8
      4.2.1. L'attribut BEH_SMO 8
      4.2.2. L'attribut BEH_PHA 8
      4.2.3. Les attribut BEH_DIET_* 9
      4.2.4. L'attribut BEH_JOB_PHA 12
      4.2.5. L'attribut BEH_MED_CHOL 12
      4.2.6. L'attribut BEH_MED_BP 13
   4.3. Les attributs de changements de facteurs de risque 13
      4.3.1. L'attribut RF_HDL 13
      4.3.2. L'attribut RF_TRIGLY 14
      4.3.3. L'attribut RF_URICACID 14
      4.3.4. L'attribut RF_LDL 15
      4.3.5. L'attribut RF_WEIGHT 15
      4.3.6. L'attribut RF_BLOODPRESS 16
      4.3.7. L'attribut RF_CHOLEST 17
      4.3.8. L'attribut RF_GLYCEMIA 17
5. Annexe 1 : Renommage des attributs 19

6. Annexe 2 25
Rappel des attributs de la table Entry 25
Rappel des attributs de la table ModifiedControl 26

1. Première modifications

Un certain nombre de modifications préliminaires ont été effectuées sur les tables de STULONG

1.1. Renommage des attributs

Pour faciliter l'analyse, nous avons renommé les attributs en utilisant des termes anglais. Le détail est donné dans Annexe 1.

1.2. Traitement des valeurs manquantes

Les valeurs manquantes et les valeurs "not stated" des tables d'origine ont été remplacées par la valeur -2.

1.3. Corrections

Nous avons effectué quelques corrections sur la table **Control** d'origine ; les modifications ont été sauvegardées dans la table **ModifiedControl.**

Dans la table **Control** d'origine, on peut observer que

- pour les patients de numéro 10102 et 10675, des contrôles consécutifs dont les dates ne respectent pas l'ordre (le contrôle numéro p a une date antérieure au contrôle numéro p-1) ; nous avons inversé ces contrôles
- pour le patient 10543, la valeur donnant le mois n'est pas renseignée pour un contrôle ; nous l'avons remplacée par la valeur 6.
- pour le patient 10448, lors d'un contrôle, les attributs indiquant le suivi de régime e la prise de médicament son incohérents ; en effet, CHG_MED_CHOLEST =70 (ne prend pas de médicament pour le cholestérol) et CHG_DIET =7 (prend des médicaments pour le cholestérol) pour ce patient. Nous avons remplacée la valeur des deux attributs par -2.

2. Objectifs

On souhaite extraire non seulement

- trouver des règles de la forme **CHG_BEH_*** $\rightarrow$ **CHG_RF_*** qui indiquent qu'un changement dans le comportement (partie gauche) a entraîné un changement dans le facteur de risque (partie droite).
- étudier les différences entre les groupes Malades/Sains, Groupe à risque/Autres Groupes, Groupe à risque RGC/Groupe à risque RGI

Nous avons construit les tables **InformationPatient** et **Changes** à partir des tables **Entry** et **ModifiedControl**

2.1. Les tables de la base de données Athero

- la table **Entry** qui comprend les même enregistrements que la table Entry d'origine et dont les attributs ont été renommés
- la table **ModifiedControl** qui comprend les enregistrements que la table Control d'origine, avec les modifications indiqués au paragraphe 1 et dont les attributs ont été renommés

La structure des tables est rappelée en Annexe2

- la table **Death** qui comprend les même enregistrements que la table Death d'origine et dont les attributs ont été renommés

- la table **Letter** qui comprend les même enregistrements que la table Letter d'origine et dont les attributs ont été renommés
- la table **InformationPatient** décrite ci-dessous
- la table **Changes** décrite ci-dessous

2.2. Structure de la table InformationPatient

Cette table est construite pour contenir des informations générales sur les patients : IDE_PATIENT_NUM, IDE_PATIENT_GRP, IDE_EDUC, DIAG_GEN

1. l'attribut IDE_PATIENT_NUM indique le numéro du patient
2. l'attribut IDE_PATIENT_GRP indique le groupe (non établi, normal, à risque, pathologique) dans lequel le patient a été classé lors de son entrée dans l'étude
3. l'attribut IDE_EDUC indique le niveau d'études du patient
4. l'attribut DIAG_GEN indique le diagnostic d'une maladie cardio-vasculaire (au moins) au cours de toute la période d'observation ou le décès du à une maladie cardio-vasculaire
5. l'attribut ALCOHOL indique la fréquence de consommation d'alcool au moment de l'entrée dans l'étude

Attribut	Type
IDE_PATIENT_NUM	Integer
IDE_PATIENT_GRP	Integer
IDE_EDUC	Integer
DIAG_GEN	Integer
ALCOHOL	Integer

2.3. Structure de la table Changes

Pour construire la table **Changes**, nous avons distingué

- les attributs caractéristiques (IDE_*)
- les attributs (BEH_*) indiquant un changement de comportement depuis le dernier contrôle
- les attributs (RF_*) indiquant un changement de facteur de risque depuis le dernier contrôle
- l'attribut DIAG indique le diagnostic d'une maladie cardio-vasculaire (au moins) au cours du contrôle

Le calcul des changements de facteurs de risque (attributs RF_*) est effectué selon les indications données dans les questions analytiques et le paragraphe *RF Definitions* (STULONG Basic Information)

La table **InformationPatient** contient 1226 lignes

La table **Changes** contient 10572 lignes pour 1226 patients différents comme dans la table **ModifiedControl**

La description (rôle, formule de calcul des valeurs, distribution des valeurs) de chaque attribut est donnée ci-dessous pour chacune des deux tables créées.

Attribut	Type	Attribut	Type
IDE_PATIENT_NUM	Integer	BEH_MED_CHOL	Integer
IDE_CONTROL_NUM	Integer	BEH_MED_BP	Integer
IDE_AGE	Integer	RF_HDL	Integer
IDE_DURATION	Integer	RF_TRIGLY	Integer
DIAG	Integer	RF_URIACID	Integer
BEH_SMO	Integer	RF_LDL	Integer
BEH_PHA	Integer	RF_WEIGHT	Integer
BEH_DIET_GEN	Integer	RF_ BLOODPRESS	Integer
BEH_DIET_FAT	Integer	RF_CHOLEST	Integer
BEH_DIET_CARBOHYDRATES	Integer	RF_GLYCEMIA	Integer
BEH_JOB_PHA	Integer		

3. La table InformationPatient

3.1. L'attribut IDE_PATIENT_GRP

Tables Entry et InformationPatient : valeurs de IDE_PATIENT_GRP

Valeur	Description	# lignes dans Entry	# patients contrôlés
-2	NA – not allocated	236	144
1	NG – normal group	168	207
2	NSG – normal studied group	40	39
3	RGI – risk group intervened	427	395
4	RGC – risk group control	432	403
5	PG – pathological group	114	38
	TOTAL	1417	1226

3.2. L'attribut IDE_EDUC

Niveau d'étude du patient : pour observer les différences de comportements entre les personnes ayant fait des études longues et les autres (ex : vérifier que les personnes ayant fait des études supérieures ont des comportements moins risqués). Les valeurs sont celles de l'attribut *Entry.SOC_EDUC_LVL*.

Table InformationPatient : valeurs de l'attribut IDE_EDUC

Valeur	Description	#patients contrôlés
-2	Not stated	14
1	Basic school	123
2	Apprentice school	343
3	Secondary school	389
4	University	357
	TOTAL	1226

3.3. L'attribut DIAG_GEN

Indique le diagnostic d'une maladie cardio-vasculaire (au moins) au cours de toute la période d'observation ou le décès du à une maladie cardio-vasculaire

Ses valeurs sont calculés à partir de l'attribut **DIAG** de la table **Changes** et de l'attribut **DEATH_CAUSE** de la table **DEATH**

NB : La table **Death3** contient 389 patients dont seulement 315 sont répertoriés dans la table **Control**

Table InformationPatient : valeurs de l'attribut DIAG_GEN

Valeur	Description		#patients contrôlés
0	No cv disease was found	pour ce patient ∀ un contrôle de numero p Changes.CHG_DIAGp =0 et le patient n'apparaît pas dans Death avec une valeur de l'attribut DEATH_CAUSE dans {5,6,7,17}	830
1	One cv disease at least was found on a control	pour ce patient ∃ un numéro de contrôle p tel que Changes.CHG_DIAGp =1 ou bien le patient apparaît dans Death avec une valeur de l'attribut DEATH_CAUSE dans {5,6,7,17}	396

3.4. L'attribut ALCOHOL

indique la fréquence de consommation d'alcool au moment de l'entrée dans l'étude

Les valeurs sont celles de l'attribut *Entry. ALC_FREQ*.

Table InformationPatient : valeurs de l'attribut ALCOHOL

Valeur	Description	#patients contrôlés
-2	Not stated	53
1	No alcohol	108
2	Occasionally	656
3	Regularly	409
	TOTAL	1226

4. La table Changes

4.1. Les attributs caractéristiques

Pour les attributs caractéristiques :

l'attribut IDE_PATIENT_NUM indique le numéro du patient

l'attribut IDE_CONTROL_NUM indique le numéro du contrôle

4.1.1. L'attribut IDE_AGE

Age du patient en années au moment du contrôle. Les intervalles sont définis de manière à avoir des cardinalités proches :

Table Changes : valeurs de l'attribut IDE_AGE

Valeur	Description	# contrôles
1	[39-52]	3393
2	[53-58]	3415
3	[59-73]	3764

Construit à partir de *Entry.SOC_YEAR_BIRTH* (année de naissance du patient) et *ModifiedControl.IDE_CONTROL_YEAR* (année du contrôle).

Table Entry : valeurs de l'attribut SOC_YEAR_BIRTH.

Valeur	# de patients	Valeur	# de patients
24	1	32	111
25	5	33	90
26	176	34	82
27	153	35	90
28	148	36	91
29	126	37	96
30	128	38	3
31	117		

4.1.2. L'attribut IDE_DURATION

Nombre de mois écoulés depuis l'entrée du patient dans l'étude.

Construit à partir de *Entry.SOC_YEAR_ENTRY, Entry.SOC_MONTH_ENTRY* , *ModifiedControl.IDE_CONTROL_YEAR* et *ModifiedControl.IDE_CONTROL_MONTH*.

4.1.3. L'attribut DIAG

Valeur	Num	ModifiedControl.QA2_DIAGx	# contrôles
No cv disease was found on this control	0	for this control $\forall$ a diagnostic x with $x \neq 0$ (Good health in general) and $x \neq 15$ (Tumor diagnostic) ModifiedControl.QA2_DIAGx = -2	9779
One cv disease at least was found on this control	1	for this control $\exists$ a diagnostic x with $x \neq 0$ and $x \neq 15$ and ModifiedControl.QA2_DIAGx $\neq$ -2	793

4.2. Les attributs de changements de comportement

4.2.1. L'attribut BEH_SMO

A partir de **Entry.SMO_INTENSITY** pour le premier contrôle et de l'attribut **Entry.CHG_DAILY_CIG** qui prend des valeurs entières de 0 à 60 ou la valeur -2 (valeur inconnue)

Premier contrôle

Valeur	Numérique	Entry.SMO_INTENSITY	CHG_DAILY_CIG (1)
Unknown	-2	-2 ou 13 (not stated) ou 6 (cigares and pipe)	*
Unknown	-2	*	-2
Stay_normal	1	1 ou 2 ou 3	< 15 et ≠ -2
Decreased	2	4 ou 5	< 15 et ≠ -2
Increased	3	1 ou 2 ou 3	>= 15
Stay_high	4	4 ou 5	>= 15

Contrôles suivants

Valeur	Numérique	CHG_DAILY_CIG (N-1)	CHG_DAILY_CIG (N)
Unknown	-2	-2	*
Unknown	-2	*	-2
Stay_normal	1	< 15 et ≠ -2	< 15 et ≠ -2
Decreased	2	>= 15	< 15 et ≠ -2
Increased	3	< 15 et ≠ -2	>= 15
Stay_high	4	>= 15	>= 15

Répartition des valeurs de **BEH_SMO**

Valeur	**Numérique**	**# contrôles**
Unknown	-2	81
Stay_normal	1	6776
Decreased	2	654
Increased	3	252
Stay_high	4	2809

4.2.2. L'attribut BEH_PHA

Premier contrôle

Valeur	Numérique	Entry.PHA_AFTER_JOB	CHG_LEISURE_PHA (1)
Unknown	-2	-2 ou 4 (not stated)	*
Unknown	-2	*	-2 ou 4 (not stated)
Stay_sits	1	1	1
Decreased_sits	2	2 ou 3	1

Increased_modest	3	1	2
Stay_modest	4	2	2
Decreased_modest	5	3	2
Increased_great	6	1 ou 2	3
Stay_great	7	3	3

Contrôles suivants

Valeur	Numérique	CHG_LEISURE_PHA (N-1)	CHG_LEISURE_PHA (N)
Unknown	-2	-2 ou 4 (not stated)	*
Unknown	-2	*	-2 ou 4 (not stated)
Stay_mostly_sits	1	1	1
Decreased_sits	2	2 ou 3	1
Increased_modest	3	1	2
Stay_modest	4	2	2
Decreased_modest	5	3	2
Increased_great	6	1 ou 2	3
Stay_great_activity	7	3	3

Répartition des valeurs de **BEH_PHA**

Valeur	Numérique	# contrôles
Unknown	-2	85
Stay_mostly_sits	1	1322
ecreased_sits	2	493
Increased_modest	3	447
Stay_modest	4	6966
Decreased_modest	5	243
Increased_great	6	282
Stay_great_activity	7	734

4.2.3. Les attribut BEH_DIET_*

Les valeurs de l'attribut **ModifiedControl.CHG_DIET** (ZMDIET) représentent déjà des changements selon 3 différents critères :

le régime alimentaire général recommandé,

une réduction des graisses,

une réduction des carbohydrates.

Nous avons créé 3 attributs pour distinguer ces régimes :

BEH_DIET_GEN, BEH_DIET_FAT, BEH_DIET_CARBOHYDRATES

L'attribut BEH_DIET_GEN

Premier contrôle

Il n'existe pas dans *Entry* d'attribut correspondant et donc la valeur associée aux premiers contrôles est '*Unknown*' (-2).

Contrôles suivants

On considère que pour *ModifiedControl.CHG_DIET*, les valeurs 3 = '*He eats sometimes according to the recommended list*' et 2 = '*He eats exactly according to the recommended list*' correspondent aux personnes qui suivent le régime général (liste de recommandations).

Les autres valeurs (1, 4, 5, 6 et 7) correspondent alors aux personnes qui ne suivent pas de régime général.

Valeur	Numérique	CHG_DIET (N-1)	CHG_DIET (N)
Unknown	-2	-2 ou 8 (not stated)	*
Unknown	-2	*	-2 ou 8 (not stated)
Unknown	-2	-------	1st control
Stay_yes	1	2 ou 3	2 ou 3
Started	2	1, 4, 5, 6 ou 7	2 ou 3
Stopped	3	2 ou 3	1, 4, 5, 6 ou 7
Stay_no	4	1, 4, 5, 6 ou 7	1, 4, 5, 6 ou 7

Répartition des valeurs de BEH_DIET_GEN

Valeur	Numérique	#contrôles
Unknown	-2	1301
Stay_yes	1	369
No_to_yes	2	375
Yes_to_no	3	296
Stay_no	4	8231

L'attribut BEH_DIET_FAT

Premier contrôle : Il n'existe pas dans *Entry* d'attribut correspondant et donc la valeur associée aux premiers contrôles est '*Unknown*' (-2).

Valeur	Numérique	CHG_DIET (N)	CHG_DIET (N+1)
Unknown	-2	-2 ou 8	*
Unknown	-2	*	-2 ou 8
Unknown	-2	-------	1st control
Stay_yes	1	4 ou 6	4 ou 6
No_to_yes	2	1, 2,3,5 ou 7	4 ou 6
Yes_to_no	3	4 ou 6	1, 2,3,5 ou 7
Stay_no	4	1, 2,3,5 ou 7	1, 2,3,5 ou 7

Répartition des valeurs de BEH_DIET_FAT

Valeur	Numérique	#contrôles
Unknown	-2	1301
Stay_yes	1	471
No_to_yes	2	568
Yes_to_no	3	557
Stay_no	4	7675

L'attribut BEH_DIET_ CARBOHYDRATES

Premier contrôle : Il n'existe pas dans *Entry* d'attribut correspondant et donc la valeur associée aux premiers contrôles est '*Unknown*' (-2).

Valeur	Numérique	CHG_DIET (N)	CHG_DIET (N+1)
Unknown	-2	-2 ou 8	*
Unknown	-2	*	-2 ou 8
Unknown	-2	-------	1st control
Stay_yes	1	5 ou 6	5 ou 6
No_to_yes	2	1, 2,3,4 ou 7	5 ou 6
Yes_to_no	3	5 ou 6	1, 2,3,4 ou 7
Stay_no	4	1, 2,3,4 ou 7	1, 2,3,4 ou 7

Répartition des valeurs de BEH_DIET_CARBOHYDRATES

Valeur	Numérique	#contrôles
Unknown	-2	1301
Stay_yes	1	363
No_to_yes	2	425
Yes_to_no	3	396
Stay_no	4	8087

4.2.4. L'attribut BEH_JOB_PHA

L'attribut *ModifiedControl.CHG_IN_JOB* représentant déjà un changement dans l'activité physique professionnelle, les transformations sont des agrégations de valeurs seulement.

Valeur	Numérique	CHG_IN_JOB (N)	#contrôles
Unknown	-2	-2 ou 50 (not stated)	43
Reduced	1	9, 29, 30, 40, 69, 79	436
Unchanged	2	20, 60, 70	10040
Increased	3	8, 48, 68, 78	53

On considère que

les valeurs 20 "no change in a job" , 60 "increased responsibility" et 70 "reduced responsibility" indiquent qu'il n'y a aucun changement dans l'activité physique au travail

les valeurs 9 "decrease of physical activities (reduced physical activities in a job", 29 "no change in a job but reduced physical activitie " 30 "leaving to partial retirement ", 40 "leaving to full retirement", 69 "increased responsibility + reduced physical activities", 79 "reduced responsibility + reduced physical activities" indiquent une diminution de l'activité physique au travail

les valeurs 8 " increase of physical activities (more physical activities in a job", 48 " leaving to full retirement and increase of physical activities", 68 "increased responsibility + increase of physical activities", 78 "reduced responsibility + increase of physical activities" indiquent une augmentation de l'activité physique au travail

4.2.5. L'attribut BEH_MED_CHOL

Valeur	Numérique	CHG_MED_CHOLEST (N)	CHG_MED_CHOLEST (N+1)
Unknown	-2	-2 ou 8	*
Unknown	-2	*	-2 ou 8
Unknown	-2	-------	1st control
Stay_yes	1	71, 72, 73, 75	71, 72, 73, 75
Started	2	70	71, 72, 73, 75
Stopped	3	71, 72, 73, 75	70
Stay_no	4	70	70

Répartition des valeurs de BEH_ MED_CHOL

Valeur	Numérique	#contrôles
Unknown	-2	7665
Stay_yes	1	93
Started	2	71
Stopped	3	13
Stay_no	4	2730

4.2.6. L'attribut BEH_MED_BP

Premier contrôle : Il n'existe pas dans *Entry* d'attribut correspondant et donc la valeur associée aux premiers contrôles est '*Unknown*' (-2).

Valeur	Num	CHG_MED_BLOODPRESS (N)	CHG_MED_ BLOODPRESS (N+1)
Unknown	-2	-2 ou =3	*
Unknown	-2	*	-2 ou =3
Unknown	-2	-------	1st control
Stay_yes	1	> 3 ou =0	> 3 ou =0
Started	2	2	> 3 ou =0
Stopped	3	> 3 ou =0	2
Stay_no	4	2	2

Une analyse plus fine pourrait distinguer des catégories (pharmaceutiques) de médicaments.

Répartition des valeurs de BEH_ MED_BP

Valeur	Numérique	#contrôles
Unknown	-2	1235
Stay_yes	1	1828
Started	2	379
Stopped	3	139
Stay_no	4	6991

4.3. Les attributs de changements de facteurs de risque

Les valeurs des seuils sont ceux de Analytical Questions et sinon (i.e. s'il n'existe pas dans ce paragraphe) de RF Definitions.

Premier contrôle : Il n'existe pas dans *Entry* d'attribut correspondant et donc la valeur associée aux premiers contrôles est '*Unknown*' (-2).

4.3.1. L'attribut RF_HDL

Valeur	Numérique	BIO_HDL_MMOL (N)	BIO_HDL_MMOL (N+1)
Unknown	-2	-2	*
Unknown	-2	*	-2
Unknown	-2	-------	1st control
Stay_normal	1	>= 1.2	>= 1.2
Increased	2	< 1.2 et ≠ -2	>= 1.2
Decreased	3	>= 1.2	< 1.2 et ≠ -2
Stay_low	4	< 1.2 et ≠ -2	< 1.2 et ≠ -2

Répartition des valeurs de RF_HDL

Valeur	Numérique	#contrôles
Unknown	-2	4841
Stay_normal	1	2526
Increased	2	890
Decreased	3	888
Stay_low	4	1427
		10572

4.3.2. L'attribut RF_TRIGLY

Valeur	Numérique	BIO_TRIGLY_MMOL (N)	BIO_TRIGLY_MMOL (N+1)
Unknown	-2	-2	*
Unknown	-2	*	-2
Unknown	-2	-------	1st control
Stay_normal	1	< 2.2 et ≠ -2	< 2.2 et ≠ -2
Decreased	2	>= 2.2	< 2.2 et ≠ -2
Inceased	3	< 2.2 et ≠ -2	>= 2.2
Stay_high	4	>= 2.2	>= 2.2

Répartition des valeurs de RF_TRIGLY

Valeur	Numérique	#contrôles
Unknown	-2	2002
Stay_normal	1	4797
Decreased	2	1034
Inceased	3	1004
Stay_high	4	1735
		10572

4.3.3. L'attribut RF_URICACID

Valeur	Numérique	BIO_URICACID (N)	BIO_URICACID (N+1)
Unknown	-2	-2	*
Unknown	-2	*	-2
Unknown	-2	-------	1st control
Stay_normal	1	<=420 et ≠ -2	<=420 et ≠ -2
Decreased	2	> 420	<=420 et ≠ -2
Inceased	3	<=420 et ≠ -2	> 420
Stay_high	4	> 420	> 420

Répartition des valeurs de CHG_RF_URICACID

Valeur	Numérique	#contrôles
Unknown	-2	10172
Stay_normal	1	332
Decreased	2	12
Inceased	3	32
Stay_high	4	24

4.3.4. L'attribut RF_LDL

Valeur	Numérique	BIO_LDL_MMOL (N)	BIO_LDL_ MMOL (N+1)
Unknown	-2	-2	*
Unknown	-2	*	-2
Unknown	-2	-------	1st control
Stay_normal	1	< 3.5 et ≠ -2	< 3.5 et ≠ -2
Decreased	2	>= 3.5	< 3.5 et ≠ -2
Inceased	3	< 3.5 et ≠ -2	>= 3.5
Stay_high	4	>= 3.5	>= 3.5

Répartition des valeurs de RF_LDL

Valeur	Numérique	#contrôles
Unknown	-2	4652
Stay_normal	1	2064
Decreased	2	697
Inceased	3	707
Stay_high	4	2452

4.3.5. L'attribut RF_WEIGHT

Valeur	Num	ENTRY.HEIGHT = h CONTROL.WEIGTH (N) = w	ENTRY.HEIGHT = h CONTROL.WEIGTH (N+1) = w
Unknown	-2	h ou w = *	h ou w = -2
Unknown	-2	h ou w = -2	h ou w = *
Stay_normal	1	W / h² <= 27	W / h² <= 27
Decreased_normal	2	27<= W / H² <30	W / H² < 27
Increased_overweight	3	W / H² < 27	27<= W / H²
Stay_overweight	4	27<= W / H² <30	27<= W / H² <30

Decreased_overweight	5	W / H² >= 30	W / H² <30
Increased_obesity	6	27<= W / H² <30	W / H² >= 30
Stay_obesity	7	W / H² >= 30	W / H² >= 30

Répartition des valeurs de RF_WEIGHT

Valeur	Numérique	#contrôles
Unknown	-2	282
Stay_normal	1	5686
Decreased_normal	2	428
Increased_overweight	3	533
Stay_overweight	4	2068
Decreased_overweight	5	211
Increased_obesity	6	275
Stay_obesity	7	1089

4.3.6. L'attribut RF_BLOODPRESS

Valeur	Num	BIO_LDL_MMOL (N)	BIO_LDL_ MMOL (N+1)
Unknown	-2	PHE_SYST = -2 ou PHE_DIAST=-2	*
Unknown	-2	*	PHE_SYST = -2 ou PHE_DIAST = -2
Stay_normal	1	(PHE_SYST < 140 et ≠ -2) et (PHE_DIAST < 90 et ≠ -2)	(PHE_SYST < 140 et ≠ -2) et (PHE_DIAST < 90 et ≠ -2)
Decreased	2	PHE_SYST >= 140 ou PHE_DIAST >= 90	PHE_SYST < 140 et (PHE_DIAST < 90 et ≠ -2)
Inceased	3	(PHE_SYST < 140 et ≠ -2) et (PHE_DIAST < 90 et ≠ -2)	PHE_SYST >= 140 ou PHE_DIAST >= 90
Stay_high	4	PHE_SYST >= 140 ou PHE_DIAST >= 90	PHE_SYST >= 140 ou PHE_DIAST >= 90

Répartition des valeurs de RF_BLOODPRESS

Valeur	Numérique	#contrôles
Unknown	-2	42
Stay_normal	1	480
Decreased	2	1180
Increased	3	4580
Stay_high	4	4290

4.3.7. L'attribut RF_CHOLEST

Valeur	Num	BIO_CHOLEST_MMOL (N)	BIO_CHOLEST_MMOL (N+1)
Unknown	-2	-2	*
Unknown	-2	*	-2
Stay_normal	1	< 6 et ≠ -2	< 6 et ≠ -2
Decreased	2	>=6	< 6 et ≠ -2
Increased	3	< 6 et ≠ -2	>= 6
Stay_high	4	>= 6	>= 6

Répartition des valeurs de RF_CHOLEST

Valeur	Numérique	#contrôles
Unknown	-2	1539
Stay_normal	1	3480
Decreased	2	1225
Increased	3	1227
Stay_high	4	3101

4.3.8. L'attribut RF_GLYCEMIA

Valeur	Num	BIO_GLYCEMIA (N)	BIO_GLYCEMIA (N+1)
Unknown	-2	-2	*
Unknown	-2	*	
Stay_normal	1	< 6.3 et ≠ -2	< 6.3 et ≠ -2
Decreased_normal	2	>= 6.3	< 6.3 et ≠ -2
Increased_disturbed	3	< 6.3 et ≠ -2	>= 6.3 et < 8
Stay_disturbed	4	>= 6.3 et < 8	>= 6.3 et < 8
Decreased_disturbed	5	>= 8	>= 6.3 et < 8
Increased_diabetes	6	< 8	>= 8
Stay_diabetes	7	>= 8	>= 8

Répartition des valeurs de RF_GLYCEMIA

Valeur	Numérique	#contrôles
Unknown	-2	10113
Unknown	1	287
Stay_normal	2	54
Decreased_normal	3	32
Increased_disturbed	4	18
Stay_disturbed	5	10
Decreased_disturbed	6	48
Increased_diabetes	7	10

5. Annexe 1 : Renommage des attributs

Table ENTRY

ICO	IDE_PATIENT_NUM
KONSKUP	IDE_PATIENT_GRP
ROKNAR	SOC_YEAR_BIRTH
ROKVSTUP	SOC_YEAR_ENTRY
MESVSTUP	SOC_MONTH_ENTRY
STAV	SOC_MARITAL_STATUS
VZDELANI	SOC_EDUC_LVL
ZODPOV	SOC_JOB_RESP
TELAKTZA	PHA_IN_JOB
AKTPOZAM	PHA_AFTER_JOB
DOPRAVA	PHA_MEAN_TRANSP
DOPRATRV	PHA_TIME_TRANSP
KOURENI	SMO_INTENSITY
DOBAKOUR	SMO_DURATION
BYVKURAK	SMO_STOP_DURATION
ALKOHOL	ALC_FREQ
PIVO7	ALC_BEER7
PIVO10	ALC_BEER10
PIVO12	ALC_BEER12
VINO	ALC_WINE
LIHOV	ALC_LIQUOR
PIVOMN	ALC_DAILY_BEER
VINOMN	ALC_DAILY_WINE
LIHMN	ALC_DAILY_LIQUOR
KAVA	SCT_DAILY_COFFEE
CAJ	SCT_DAILY_TEA
CUKR	SCT_DAILY_SUGAR
IM	PER_IM
IML	PER_IM_MED
HT	PER_HT
HTD	PER_HT_DIET
HTL	PER_HT_MED
ICT	PER_ICT
ICTL	PER_ICT_MED
DIABET	PER_DIAB
DIABD	PER_DIAB_DIET

DIABL PER_DIAB_MED
HYPLIP PER_HYPERLIP
HYPLD PER_HYPERLIP_DIET
HYPLL PER_HYPERLIP_MED
IMTRV PER_IM_DURATION
HTTRV PER_HT_DURATION
ICTTRV PER_ICT_DURATION
DIABTRV PER_DIAB_DURATION
HYPLTRV PER_HYPERLIP_DURATION
BOLHR QA2_CHEST_PAIN
BOLDK QA2_LIMB_PAIN
DUSNOST QA2_ASTHMA
VYSKA PHE_HEIGHT
VAHA PHE_WEIGHT
SYST1 PHE_SYST1
DIAST1 PHE_DIAST1
SYST2 PHE_SYST2
DIAST2 PHE_DIAST2
TRIC PHE_TRICEPS
SUBSC PHE_SUBSC
CHLST BIO_CHOLEST
TRIGL BIO_TRIGLY
MOC BIO_URINE
RARISK RSK_FAMILY_ANAMN
OBEZRISK RSK_OBESITY
KOURRISK RSK_SMOKING
HTRISK RSK_HT
CHOLRISK RSK_INC_CHOLEST

Table CONTROL

ICO IDE_PATIENT_NUM REAL
ROKVYS IDE_CONTROL_YEAR REAL
MESVYS IDE_CONTROL_MONTH REAL
PORADK IDE_CONTROL_NUM REAL
ZMCHARZA CHG_IN_JOB REAL
ZMTELAKT CHG_PHA REAL
AKTPOZAM CHG_LEISURE_PHA REAL
ZMDIET CHG_DIET REAL
LEKCHOL CHG_MED_CHOLEST REAL

LEKTLAK CHG_MED_BLOODPRESS REAL
ZMKOUR CHG_SMOKING REAL
POCCIG CHG_DAILY_CIG REAL
PRACNES SCK_HASBEEN_LEAVE REAL
SRDCE SCK_HEARTDIS_FREQ REAL
HYPERT SCK_HT_FREQ REAL
CEVMOZ SCK_CEREBRO_FREQ REAL
DIAB SCK_DIAB_FREQ REAL
JINAONE SCK_OTHER_FREQ REAL
BOLHR QA2_CHEST_PAIN REAL
BOLDK QA2_LIMB_PAIN REAL
DUSN QA2_DYSPNEA REAL
HODNSK QA2_CURRENT_GRP REAL
HODN0 QA2_DIAG_0 REAL
ROK0 QA2_YEAR_0 REAL
HODN1 QA2_DIAG_1 REAL
ROK1 QA2_YEAR_1 REAL
HODN2 QA2_DIAG_2 REAL
ROK2 QA2_YEAR_2 REAL
HODN3 QA2_DIAG_3 REAL
ROK3 QA2_YEAR_3 REAL
HODN4 QA2_DIAG_4 REAL
ROK4 QA2_YEAR_4 REAL
HODN11 QA2_DIAG_11 REAL
ROK11 QA2_YEAR_11 REAL
HODN12 QA2_DIAG_12 REAL
ROK12 QA2_YEAR_12 REAL
HODN13 QA2_DIAG_13 REAL
ROK13 QA2_YEAR_13 REAL
HODN14 QA2_DIAG_14 REAL
ROK14 QA2_YEAR_14 REAL
HODN15 QA2_DIAG_15 REAL
ROK15 QA2_YEAR_15 REAL
HODN21 QA2_DIAG_21 REAL
ROK21 QA2_YEAR_21 REAL
HODN23 QA2_DIAG_23 REAL
ROK23 QA2_YEAR_23 REAL
HMOT PHE_WEIGTH REAL
SYST PHE_SYST REAL

DIAST PHE_DIAST REAL
TRIC PHE_TRICEPS REAL
SUBSC PHE_SUBSC REAL
HYPERSD PHE_SYSTDIAST_HT REAL
HYPERS PHE_SYST_HT REAL
HYPERD PHE_DIAST_HT REAL
HYPCHL PHE_HYPERCHOLEST REAL
HYPTGL PHE_HYPERTRIGLY REAL
CHLST BIO_CHOLEST_MMOLL REAL
CHLSTMG BIO_CHOLEST REAL
TRIGL BIO_TRIGLY_MMOLL REAL
TRIGLMG BIO_TRIGLY REAL
HDL BIO_HDL_MMOLL REAL
HDLMG BIO_HDL REAL
MOC BIO_URINE REAL
GLYKEMIE BIO_GLYCEMIA REAL
KYSMOC BIO_URICACID REAL
LDL BIO_LDL REAL

Table DEATH

ICO IDE_PATIENT_NUM
DENUMR DEATH_DAY
MESUMR DEATH_MONTH
ROKUMR DEATH_YEAR
PRICUMR DEATH_CAUSE

Table LETTER

ICO IDE_PATIENT_NUM
MESDOT IDE_YEAR_QUEST
ROKDOT IDE_MONTH_QUEST
LEKCHOL HEA_DRUGS_CHOLEST
LEKTK HEA_DRUGS_BLOODPRESS
NEMOC1 HEA_DISEASE_1
ROK1 HEA_YEAR_1
NEMOC2 HEA_DISEASE_2
ROK2 HEA_YEAR_2
NEMOC3 HEA_DISEASE_3
ROK3 HEA_YEAR_3
NEMOC4 HEA_DISEASE_4

ROK4 HEA_YEAR_4
NEMOC5 HEA_DISEASE_5
ROK5 HEA_YEAR_5
HYPTK ANSDIS_HT_DIAG
ROKHYPTK ANSDIS_YEAR_HT_DIAG
HYPLP ANSDIS_HYPERLIP_DIAG
ROKHYPLP ANSDIS_YEAR_HYPERLIP_DIAG
CUKROVKA ANSDIS_DIAB_DIAG
ROKCUKR ANSDIS_YEAR_DIAB_DIAG
CUKRTAB ANSDIS_DIAB_MED_TABLETS
ODCUTAB ANSDIS_YEAR_DIAB_TABLETS_START
DOCUTAB ANSDIS_YEAR_DIAB_TABLETS_END
CUKRINS ANSDIS_DIAB_MED_INSULIN
ODCUINS ANSDIS_DIAB_INSULIN_START
DOCUINS ANSDIS_DIAB_INSULIN_END
AP ANSDIS_AP_MED
SI ANSDIS_IM_DIAG
ROKSI ANSDIS_YEAR_IM_DIAG
MM ANSDIS_ICT_DIAG
ROKMM ANSDIS_YEAR_ICT_DIAG
BDK ANSDIS_LIMB_PAIN_MED
ROKBDK1 ANSDIS_YEAR_LIMB_PAIN_MED
DUSNOST ANSDYS_DYSPNEA
ROKDUS ANSDYS_YEAR_DYSPNEA
DUSCHUZE ANSDYS_DYSPNEA_ASCENDING
DUSBEH ANSDYS_DYSPNEA_RUNNING
DUSROVIN ANSDYS_DYSPNEA_WALKING
DUKLID ANSDYS_DYSPNEA_RESTING
DUSNOC ANSDYS_DYSPNEA_NOCTURNAL
KURAK ANSSMO_SMOKER
KURAKBYV ANSSMO_FORMER_SMOKER
ROKPRKOUR ANSSMO_STOP_DURATION
KOURLET ANSSMO_SMOKE_DURATION
CIGDENMIN ANSSMO_FORMER_DAILY_CIG
CIGDEN ANSSMO_CURRENT_DAILY_CIG
KOURODLET ANSSMO_AGE_SMOKE_START
DYMKA ANSSMO_CURRENT_PIPECIGAR
PASED ANSPHA_LOW
PAMIRNA ANSPHA_MEDIUM

PAVELKA	ANSPHA_HIGH
PAPRED10	ANSPHA_10YEARS_AGO
DIETA	ANSDIE_DIET
DIECUOD	ANSDIE_YEAR_SUGAR
DIETUKOD	ANSDIE_YEAR_LIPID
JINADIE	ANSDIE_OTHER
JINADIEOD	ANSDIE_YEAR_OTHER
VAHA	ANSWEI_CURRENT
VAHAPRED10	ANSWEI_10YEARS_AGO

6. Annexe 2

Rappel des attributs de la table Entry

Attribut	Type	Attribut	Type
IDE_PATIENT_NUM	integer	PER_ICT_MED	nominal
IDE_PATIENT_GRP	nominal	PER_DIAB	nominal
SOC_YEAR_BIRTH	integer	PER_DIAB_DIET	boolean
SOC_YEAR_ENTRY	integer	PER_DIAB_MED	nominal
SOC_MONTH_ENTRY	integer	PER_HYPERLIP	nominal
SOC_MARITAL_STATUS	nominal	PER_HYPERLIP_DIET	nominal
SOC_EDUC_LVL	nominal	PER_HYPERLIP_MED	nominal
SOC_JOB_RESP	nominal	PER_IM_DURATION	nominal
PHA_IN_JOB	nominal	PER_HT_DURATION	nominal
PHA_AFTER_JOB	nominal	PER_ICT_DURATION	nominal
PHA_MEAN_TRANSP	nominal	PER_DIAB_DURATION	nominal
PHA_TIME_TRANSP	nominal	PER_HYPERLIP_DURATION	nominal
SMO_INTENSITY	nominal	QA2_CHEST_PAIN	nominal
SMO_DURATION	nominal	QA2_LIMB_PAIN	nominal
SMO_STOP_DURATION	nominal	QA2_ASTHMA	nominal
ALC_FREQ	nominal	PHE_HEIGHT	integer
ALC_BEER	nominal	PHE_WEIGHT	integer
ALC_WINE	boolean	PHE_SYST1	integer
ALC_LIQUOR	boolean	PHE_DIAST1	integer
ALC_DAILY_BEER	nominal	PHE_SYST2	integer
ALC_DAILY_WINE	nominal	PHE_DIAST2	integer
ALC_DAILY_LIQUOR	nominal	PHE_TRICEPS	integer
SCT_DAILY_COFFEE	nominal	PHE_SUBSC	integer
SCT_DAILY_TEA	nominal	BIO_CHOLEST	integer
SCT_DAILY_SUGAR	nominal	BIO_TRIGLY	integer
PER_IM	nominal	BIO_URINE	nominal
PER_IM_MED	nominal	RSK_FAMILY_ANAMN	nominal
PER_HT	nominal	RSK_OBESITY	nominal
PER_HT_DIET	boolean	RSK_SMOKING	nominal
PER_HT_MED	nominal	RSK_HT	nominal
PER_ICT	nominal	RSK_INC_CHOLEST	nominal

Rappel des attributs de la table ModifiedControl

Attribut	Type	Attribut	Type
IDE_PATIENT_NUM	numeric	QA2_DIAG_12	Nominal
IDE_CONTROL_YEAR	numeric	QA2_YEAR_12	Numeric
IDE_CONTROL_MONTH	numeric	QA2_DIAG_13	Nominal
IDE_CONTROL_NUM	numeric	QA2_YEAR_13	Numeric
CHG_IN_JOB	nominal	QA2_DIAG_14	Nominal
CHG_PHA	nominal	QA2_YEAR_14	Numeric
CHG_LEISURE_PHA	nominal	QA2_DIAG_15	Nominal
CHG_DIET	nominal	QA2_YEAR_15	Numeric
CHG_MED_CHOLEST	nominal	QA2_DIAG_21	nominal
CHG_MED_BLOODPRESS	nominal	QA2_YEAR_21	numeric
CHG_SMOKING	nominal	QA2_DIAG_23	nominal
CHG_DAILY_CIG	nominal	QA2_YEAR_23	numeric
SCK_HASBEEN_LEAVE	nominal	PHE_WEIGTH	numeric
SCK_HEARTDIS_FREQ	nominal	PHE_SYST	numeric
SCK_HT_FREQ	nominal	PHE_DIAST	numeric
SCK_CEREBRO_FREQ	nominal	PHE_TRICEPS	numeric
SCK_DIAB_FREQ	nominal	PHE_SUBSC	numeric
SCK_OTHER_FREQ	nominal	PHE_SYSTDIAST_HT	nominal
QA2_CHEST_PAIN	nominal	PHE_SYST_HT	nominal
QA2_LIMB_PAIN	nominal	PHE_DIAST_HT	nominal
QA2_DYSPNEA	nominal	PHE_HYPERCHOLEST	nominal
QA2_CURRENT_GRP	nominal	PHE_HYPERTRIGLY	nominal
QA2_DIAG_0	nominal	BIO_CHOLEST_MMOLL	numeric
QA2_YEAR_0	numeric	BIO_CHOLEST	numeric
QA2_DIAG_1	nominal	BIO_TRIGLY_MMOLL	numeric
QA2_YEAR_1	numeric	BIO_TRIGLY	numeric
QA2_DIAG_2	nominal	BIO_HDL_MMOLL	numeric
QA2_YEAR_2	numeric	BIO_HDL	numeric
QA2_DIAG_3	nominal	BIO_URINE	nominal
QA2_YEAR_3	numeric	BIO_GLYCEMIA	numeric
QA2_DIAG_4	nominal	BIO_URICACID	numeric
QA2_YEAR_4	numeric	BIO_LDL	numeric
QA2_DIAG_11	nominal		
QA2_YEAR_11	numeric		