

## RAPPORT DEVOIR MAISON BIG DATA

# Dépistage de la Gonorrhée : Analyse des Facteurs de Risque et Caractérisation des Groupes à Risque

Réalisé par

MOIELWAY NGARTI Israël & Philia EZIN

# Introduction

Les maladies transmissibles sexuellement (MTS) représentent un enjeu majeur de santé publique en raison de leur impact sur la population et des préjugés qui les entourent. Bien que les avancées médicales aient permis des dépistages efficaces, la réussite de ces programmes repose largement sur l'accès aux populations à risque et la mise en place des stratégies adaptées.

Le dépistage des MTS, et plus particulièrement de la gonorrhée, doit ainsi se concentrer sur des actions ciblées visant à sensibiliser et à toucher des groupes spécifiques. Ce travail porte sur l'analyse de données recueillies par des médecins dans le cadre d'un programme de dépistage de la gonorrhée, permettant d'étudier les facteurs de risque associés à cette maladie.

L'objectif est d'explorer la manière dont les données sur l'âge, le sexe, l'orientation sexuelle, le nombre de partenaires sexuels et les antécédents de MTS peuvent contribuer à mieux cibler les groupes à risque et améliorer les stratégies de prévention.

## ETUDE ANALYTIQUE

### 0.1 Description des données

1. **IDENTIFICATION** : Numéro d'identification du patient (Variable numérique, identifiant unique.)
  2. **SEXE** : 1 = homme, 2 = femme (Variable catégorielle.)
  3. **ÉTAT CIVIL** : 1 = célibataire, 2 = marié, 3 = séparé ou divorcé, 4 = veuf, 5 = pas de réponse (Variable catégorielle.)
  4. **ÂGE** : Âge du patient (Variable numérique.)
  5. **ORIENTATION SEXUELLE** : 1 = homosexuel(le), 2 = hétérosexuel(le), (Variable catégorielle.)
  6. **MTS ANTÉRIEURES** : 1 = non, 2 = oui (Variable catégorielle indiquant si le patient a déjà eu une MTS).
  7. **NOMBRES DE MTS ANT.** : Nombre de maladies transmissibles sexuellement contractées avant cette visite (Variable numérique.)
  8. **RAISON DE LA VISITE** : 1 = symptôme, 2 = contact (avec une personne infectée), 3 = dépistage, 4 = contrôle (suivi médical), 5 = autre (Variable catégorielle.)
  9. **NOMBRE DE PARTENAIRES** : Nombre de partenaires sexuels (Variable numérique.)
  10. **HISTOIRE DE CONTACT** : 0 = pas de relation avec un(e) partenaire contaminé(e), 1 = relation avec partenaire(s) contaminé(s) par une MTS (Variable catégorielle.)
  11. **CULTURE** : Résultat de la culture visant à déceler la présence de gonorrhée dans différents sites : 0 = négatif, 1 = gorge (positif), 2 = col ou urètre (positif), 3 = anus (positif), 4 = gorge et urètre ou col (positif), 5 = gorge et anus (positif), 6 = urètre ou col et anus (positif), 7 = les 3 sites (positif) (Variable catégorielle indiquant les sites où la bactérie est détectée.)
  12. **DIAGNOSTIQUE** : 0 = pas de gonorrhée, 1 = gonorrhée (Variable catégorielle binaire indiquant la conclusion diagnostique.)
- Observations : 3144 individus

TABLE 1 – Statistiques descriptives

	AGE	NB_PART	NB_MTS
<b>count</b>	3144.00	3144.00	3144.00
<b>mean</b>	28.58	5.20	1.21
<b>std</b>	8.40	15.02	2.56
<b>min</b>	14.00	0.00	0.00
<b>25%</b>	23.00	1.00	0.00
<b>50%</b>	27.00	2.00	1.00
<b>75%</b>	37.00	3.00	1.00
<b>max</b>	99.00	99.00	65.00

Le tableau ci-dessus présente les statistiques descriptives des variables AGE, NB\_PART et NB\_MTS qui sont des variables quantitatives continues. Ces variables permettent un calcul pertinent de la moyenne, de la dispersion et des quartiles.

L'âge moyen des patients est de 28,58 ans, avec un écart-type de 8,40, ce qui montre une population plutôt jeune, mais avec une certaine dispersion (de 14 à 99 ans).

Le nombre moyen de partenaires sexuels au cours du dernier mois est de 5,20, mais la médiane est de 2, ce qui indique une distribution légèrement asymétrique (quelques patients avec un nombre très élevé de partenaires, jusqu'à 99, tirent la moyenne vers le haut). On peut suggérer que le profil dominant est de 2 ou 3 partenaires.

Le nombre moyen de maladies sexuellement transmissibles antérieures (NB\_MTS) est de 1,21, avec un maximum de 65. Là aussi, la médiane est plus basse, suggérant qu'une minorité de patients cumule un grand nombre d'antécédents.

Ces valeurs très élevées montrent qu'il y a les valeurs manquantes qui sont remplacées par 99.

## 0.2 Traitement des valeurs manquantes

Dans notre base de données il y a les valeurs manquantes qui sont identifiées par des champs où les valeurs sont (blancs, «9», «99», etc.).

- *Méthode utilisée pour le traitement de ces valeurs.*

Pour traiter ces valeurs manquantes, on a fait des graphiques et calculé des pourcentages afin de mieux observer la proportion de ces valeurs manquantes dans chaque variable.

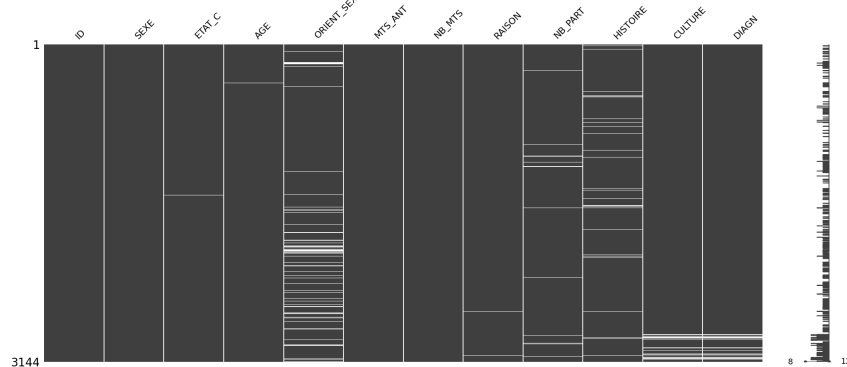


FIGURE 1 – Graphique de dispersion des valeurs manquantes

Les barres noires pleines montrent qu'il n'y a pas de valeurs manquantes dans ces variables tant dis que celle teintes de blanc montrent la proportion des valeurs manquantes.

	Missing Values	Missing Percent
<b>ID</b>	0	0.000000
<b>SEXE</b>	0	0.000000
<b>ETAT_C</b>	4	0.127226
<b>AGE</b>	6	0.190840
<b>ORIENT_SEX</b>	257	8.174300
<b>MTS_ANT</b>	0	0.000000
<b>NB_MTS</b>	0	0.000000
<b>RAISON</b>	2	0.063613
<b>NB_PART</b>	66	2.099237
<b>HISTOIRE</b>	107	3.403308
<b>CULTURE</b>	73	2.321883
<b>DIAGN</b>	73	2.321883

FIGURE 2 – Dispersion des valeurs manquantes en pourcentage

Nous constatons que le taux de valeurs manquantes varie selon les variables, il est très faible dans certains cas, considérable pour d'autres, et modéré pour le reste.

Ces données manquantes peuvent être dues à des questions sensibles (orientation sexuelle, comportements sexuels) ou à des données médicales non renseignées. Les valeurs manquantes observées pour HISTOIRE, CULTURE et DIAGN peuvent s'expliquer par des données non renseignées ou perdues, des examens médicaux non réalisés ou des patients ne connaissant pas leur historique médical. Elles devront donc être prises en compte dans la suite de l'analyse soit par suppression, soit par imputation, selon le type de variable et l'importance de la perte d'information.

Après analyse, nous avons adopté la stratégie suivante pour le traitement :

- ETAT\_C (état civil), AGE (âge), RAISON (raison) : compte tenu de leur très faible pourcentage de valeurs manquantes, nous avons choisi de supprimer ces observations incomplètes.
- ORIENTATION SEXUELLE : nous avons créé une troisième catégorie pour regrouper les données manquantes.
- Autres variables : nous avons décidé d'imputer les valeurs manquantes par le mode afin de préserver autant que possible l'information disponible.

Après le traitement des valeurs manquantes, nous procédons à l'analyse des variables, nous allons essayer de voir la distribution des catégories dans chaque variable, la relation qui existe entre certaines variables et la variable diagnostique.

Pour les variables catégorielles, nous générons des diagrammes en barres qui illustrent les distributions.

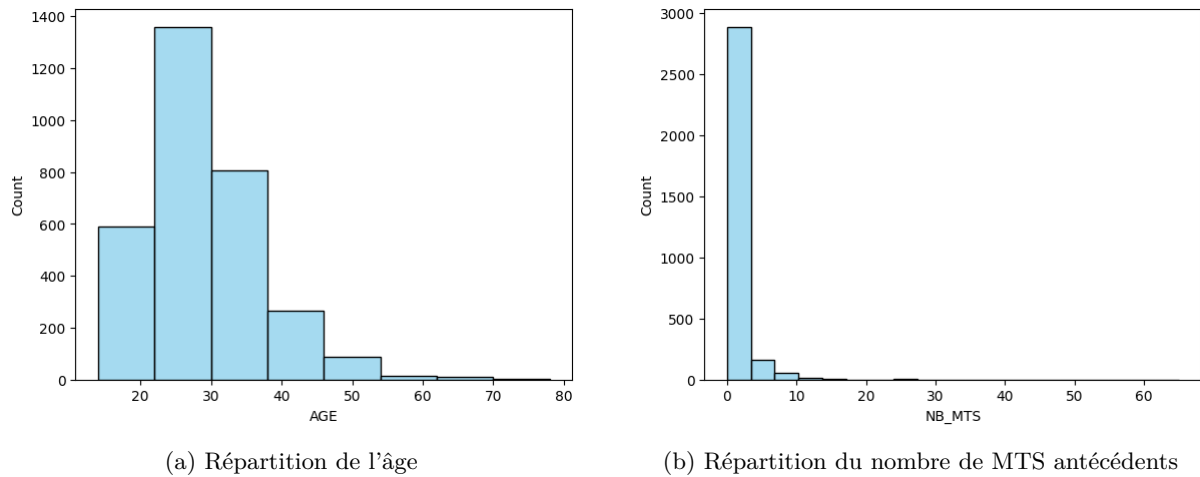


FIGURE 3 – Distributions de l'âge et du nombre de MTS antécédents

- Pour la Figure 3a : Ce graphique présente un histogramme montrant la répartition de l'âge avec l'âge en abscisses et le nombres d'individus en ordonnées. De cet histogramme, on peut remarquer que la majorité des individus se situe entre 20 et 30 ans, avec un pic autour de 25 ans. Ensuite il y a une forte diminution après 30 ans, et très peu de personnes au-delà de 50 ans.
- Pour la Figure 3b : L'axe x représente le nombre de maladies transmises sexuellement antécédents (NB\_MTS) et l'axe y (Count) indique le nombre de personnes concernées pour chaque valeur de NB\_MTS. L'écrasante majorité des individus ont eu 0 ou très peu de MTS : on voit un pic très haut à 0, puis une chute rapide. Quelques cas isolés vont jusqu'à plus de 20 ou 30 antécédents, mais ce sont des valeurs extrêmes.

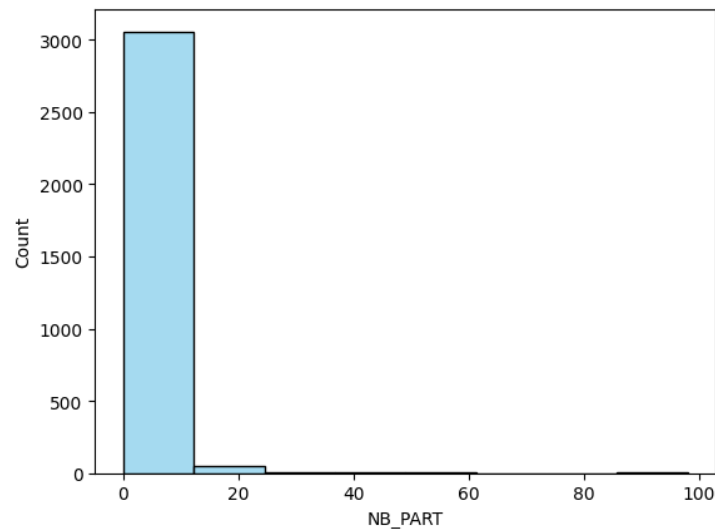


FIGURE 4 – Répartition du nombre de partenaire

La figure 1.5 montre la répartition du nombre de partenaire. On voit un très fort regroupement entre 0 et 10 partenaires, avec un pic à gauche. La distribution est fortement asymétrique à droite ( quelques valeurs s'étendent jusqu'à 100 partenaires).

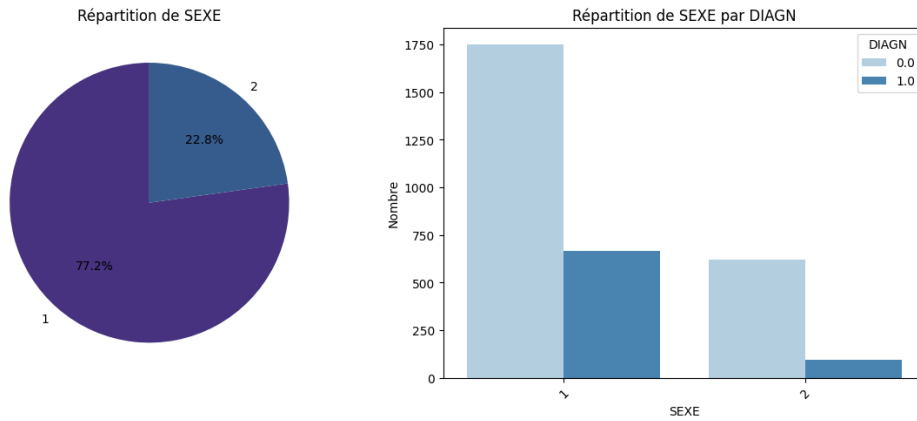


FIGURE 5 – Distribution du sexe en fonction du diagnostic

Le camembert à gauche montre que la majorité des individus sont des hommes et la minorité, des femmes. Parmi cette majorité d’hommes, un nombre significatif a reçu un diagnostic positif pour la gonorrhée. Concernant les femmes, très peu de cas de gonorrhée ( $DIAGN = 1$ ) sont recensés parmi elles. Cela pourrait indiquer une prévalence plus faible de la gonorrhée chez les femmes et que la gonorrhée touche majoritairement les hommes.

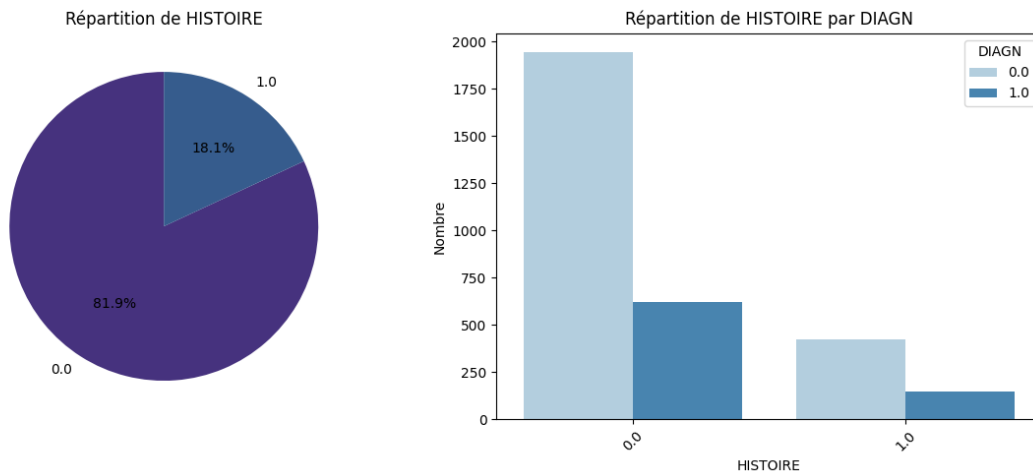


FIGURE 6 – Distribution de l’histoire en fonction du diagnostic

Du graphique ci dessus, les individus n’ayant pas de contact connu avec un partenaire contaminé ( $HISTOIRE = 0$ ) sont la majorité des cas dans l’échantillon. On observe malgré tout un nombre significatif de diagnostics positifs ( $DIAGN = 1$ ) pour la gonorrhée dans ce groupe. Ceux ayant un contact avec un partenaire contaminé ( $HISTOIRE = 1$ ) sont moins représenté globalement. Mais une proportion notable de diagnostics positifs y est aussi observée.

On pourrait donc dire que les individus ayant eu un contact avec un partenaire MTS sont plus à risque de contracter la gonorrhée.

Ce graphique illustre comment la répartition des cas de gonorrhée varie en fonction de l’état civil. On constate que les célibataires sont le groupe le plus représenté et que la majorité des diagnostics de gonorrhée se trouvent dans ce groupe. Les mariés ( $ETAT\_C = 2$ ) représentent une minorité du total, avec un nombre modéré de cas positifs. Cela peut refléter une moindre exposition au risque, en lien avec des comportements plus stables sur le plan relationnel. Les personnes séparées/divorcées sont très peu représentées. Quelques cas de gonorrhée sont tout de même présents, probablement en lien avec un retour à des pratiques sexuelles à risque post-séparation. Les veufs sont pratiquement absents de l’échantillon, avec une quasi-absence de diagnostic positif. On retient que la majorité des cas de gonorrhée sont observés chez les célibataires, probablement en lien avec une plus forte activité sexuelle ou une multiplicité de partenaires.

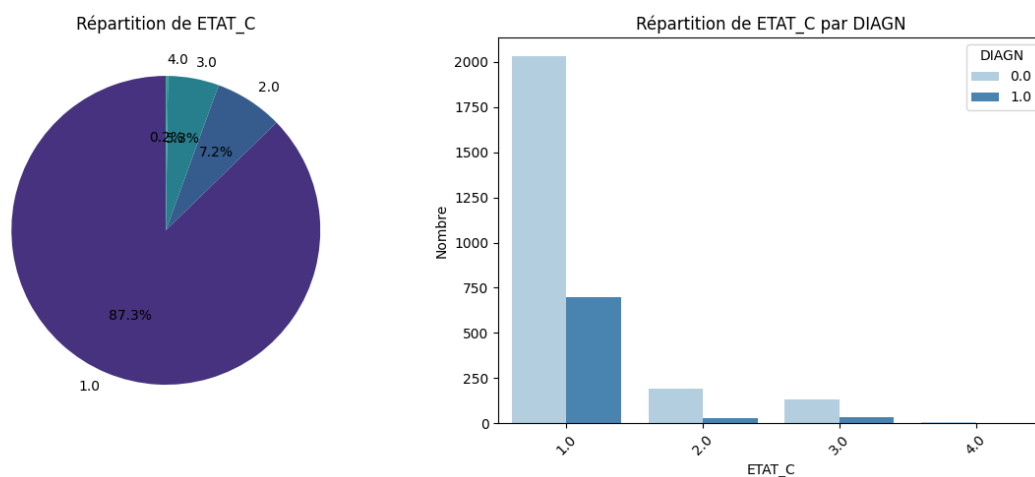


FIGURE 7 – Distribution de l'état civil en fonction du diagnostic

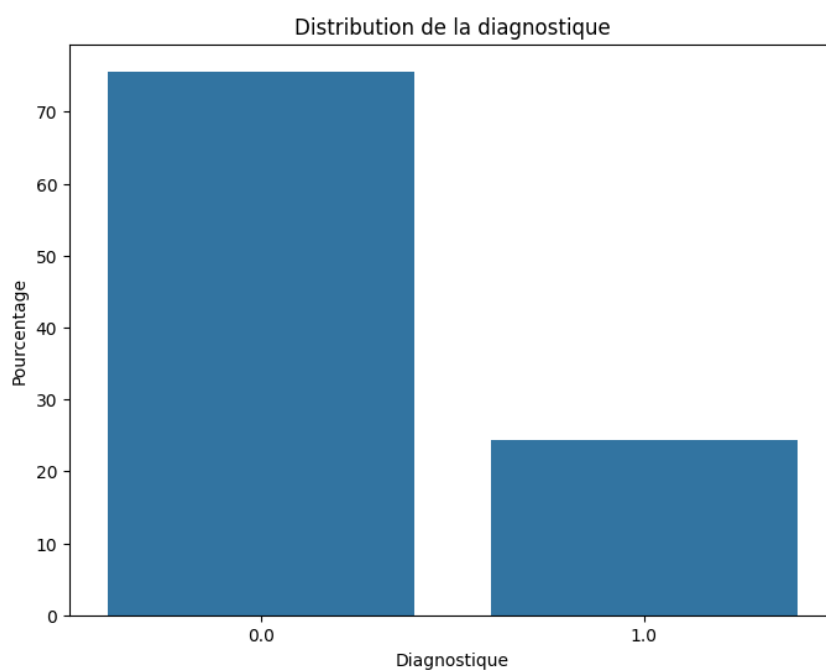


FIGURE 8 – Distribution Diagnostique

Dans cette étude, le nombre d'individus diagnostiqués comme n'ayant pas la gonorrhée est nettement supérieur à celui des individus positifs, ce qui révèle un déséquilibre de classes assez conséquent.

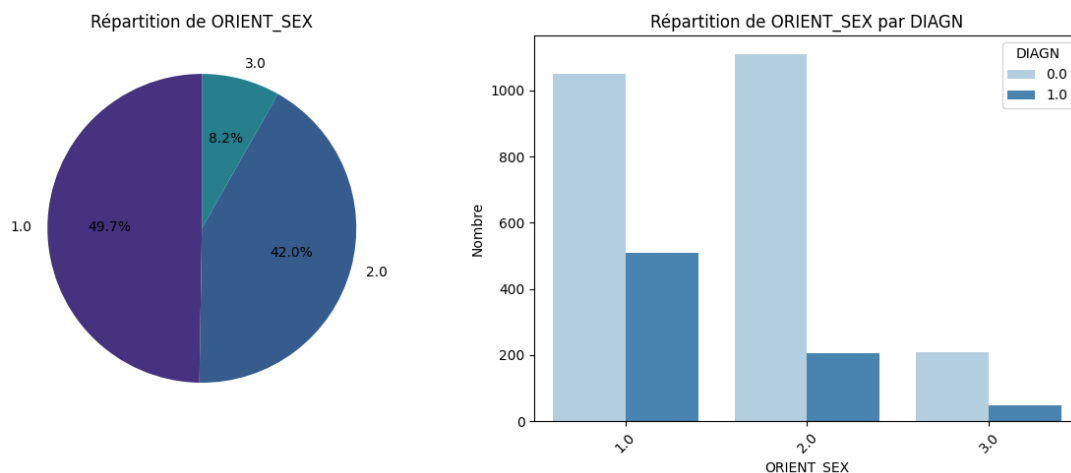


FIGURE 9 – Distribution de l'orientation sexuelle en fonction du diagnostic

On observe une surreprésentation des personnes homosexuelles dans l'échantillon, ce qui pourrait refléter une plus grande propension à consulter dans les cliniques MTS. Chez les homosexuels, un nombre important ont un diagnostic positif (DIAGN = 1), bien que la majorité reste négative. Chez les hétérosexuels (2.0), moins de cas positifs sont observés, malgré une proportion globale élevée dans l'échantillon. Les autres (3.0) représentent un faible effectif, mais quelques cas positifs existent.

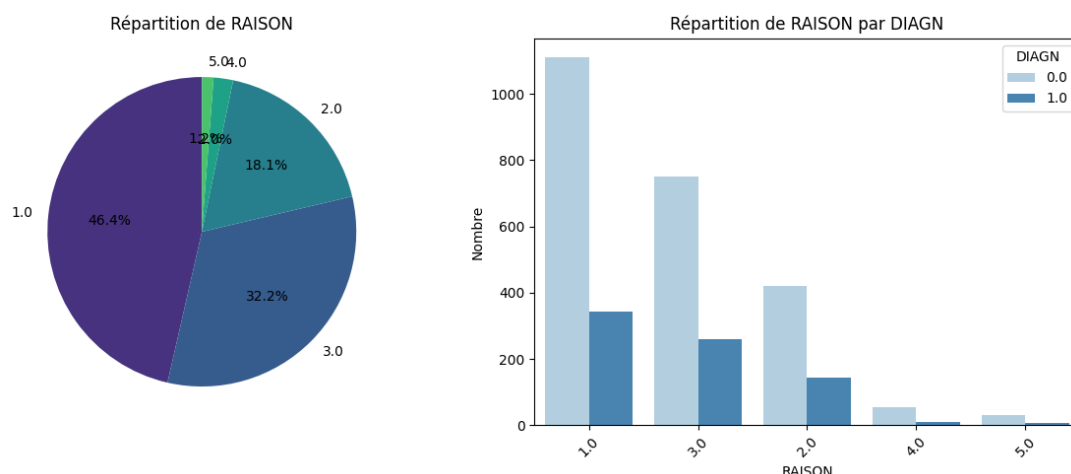


FIGURE 10 – Distribution de la raison en fonction du diagnostic

Les résultats observés dans la figure 10 montrent que les consultations pour symptôme (RAISON=1) représentent près de la moitié des cas. Cela reflète que les patients consultent principalement lorsqu'ils ressentent des signes cliniques. Cependant, on remarque que ceux qui sont positifs à la gonorrhée sont proportionnellement plus élevés dans les personnes venues pour contact et dépistage. Cela peut s'expliquer par le fait que les contacts d'individus infectés ont une probabilité plus élevée d'être eux-mêmes porteurs. Les personnes venant se faire dépister sont peut-être plus à risque (comportements à risque, partenaires multiples). Cela suggère que le motif de consultation est lié à la probabilité d'infection, et pourrait donc être une variable explicative importante dans un éventuel modèle de régression visant à prédire la gonorrhée.

Le graphique circulaire montre que la très grande majorité des patients présente un résultat de culture négatif, tandis que la minorité présente une culture positive uniquement à la gorge. On observe que la majorité des individus dont le résultat de la culture est négatif présentent également un diagnostic négatif, ce qui est logique dans la mesure où l'absence de prélèvements se traduit naturellement par l'absence de diagnostic. Ces résultats suggèrent que la variable «CULTURE» pourrait biaiser notre modèle de prédiction. Par conséquent, il serait pertinent d'envisager de la retirer ou de la retravailler lors de l'optimisation du modèle.



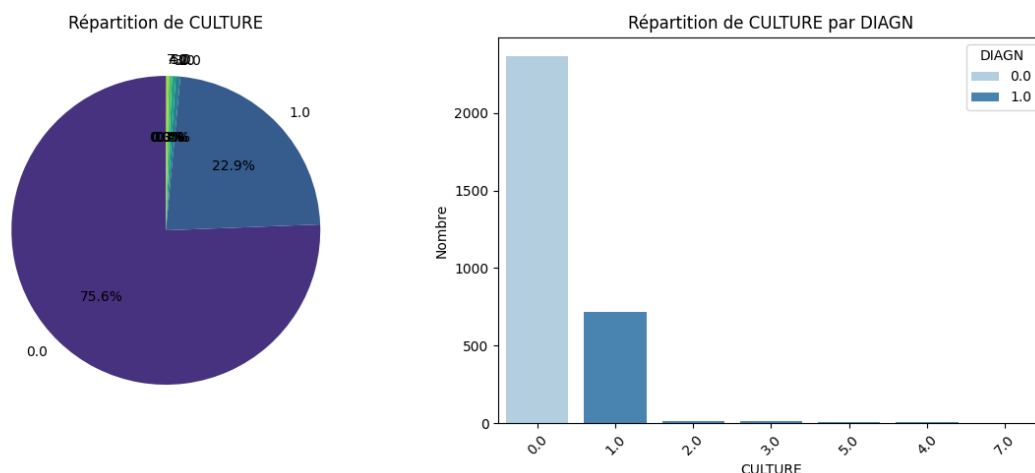


FIGURE 11 – Répartition de culture par diagnostic

## Méthodologie d'analyse des groupes à risque

### 0.3 Catégorisation des variables

Afin d'avoir toutes les variables en catégorielle, on catégorise les variables numériques. - Catégories d'âge : moins de 30 ans et 30 ans et plus

- Catégories pour le nombre de MTS antérieures (déjà et pas encore)

- Nombre de partenaires (peu actif et très actif) : on considère les individus avec un nombre de partenaire en dessous de la moyenne comme peu actif et au dessus comme très actif.

Après la catégorisation de ces variables on fait l'encodage avec le one-hot-encoding.

Par la suite on retire les colonnes (AGE, Nombre de MTS antérieur, Nombre de partenaire) qui ont été catégorisé pour éviter la multicollinéarité.

### 0.4 Outils et techniques statistiques utilisés

Afin d'étudier la relation entre certaines variables et la variable dépendante, on fait des tableaux croisés dynamiques et le test de chi-deux.

TABLE 2 – Répartition du diagnostic par groupe d'âge

AGE_GROUP	0.0	1.0	All	Prévalence (%)
<30	1437	512	1949	26.27
≥ 30	931	252	1183	21.30
All	2368	764	3132	24.39

La probabilité d'être diagnostiqué gonorrhée est de 26,24% pour les individus de moins 30 ans et 21,27% pour ceux de plus de 30 ans.

TABLE 3 – Distribution du diagnostic par groupe d'âge et groupe de partenaires

AGE_GROUP	PARTNER_GROUP	0.0	1.0	All	Prévalence (%)
<30	Peu actifs	1030	275	1305	21.07
<30	Très actifs	407	237	644	36.80
≥ 30	Peu actifs	612	120	732	16.39
≥ 30	Très actifs	319	132	451	29.27
All		2368	764	3132	24.39

Les individus de moins de 30 ans et très actifs ont plus de probabilité d'être diagnostiqués gonorrhée.

TABLE 4 – Distribution du diagnostique par groupe de partenaires

PARTNER_GROUP	0.0	1.0	All	Prévalence (%)
Peu actifs	1642	395	2037	19.39
Très actifs	726	369	1095	33.70
<b>All</b>	2368	764	3132	24.39

Les individus très actifs ont 33,70% et les individus Peu actifs ont 19,39% de de probilité d'être diagnostiqués gonorrhée.

TABLE 5 – Distribution du diagnostic selon l'historique de MTS (MTS\_ANT)

MTS_ANT	0.0	1.0	All	Prévalence (%)
1	1159	349	1508	23.14
2	1209	415	1624	25.55
<b>All</b>	2368	764	3132	24.39

Les individus ayant un antécédent de MTS ont 25,55% et ceux n'ayant pas un antécédent ont 23,14% de probabilité d'être diagnostiqués gonorrhée.

-Le test de Khi-deux

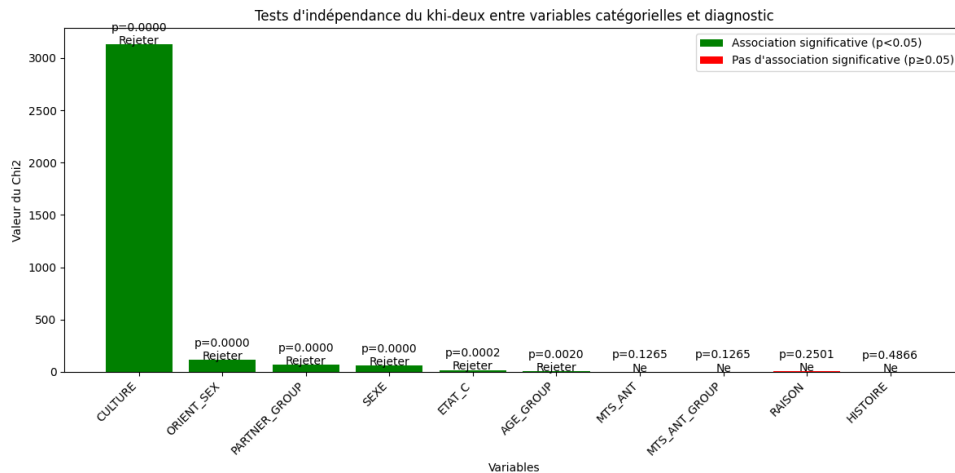


FIGURE 12 – Résultat du test de khi-deux

Les résultats indiquent quelles variables sont significativement associées au diagnostic ( $p < 0,05$ ) et lesquelles ne le sont pas ( $p \geq 0,05$ ). Autrement dit :

Variables non indépendantes du diagnostic ( $p < 0,05$ ) : CULTURE, ORIENT\_SEX, PARTNER\_GROUP, SEXE, ETAT\_C, AGE\_GROUP.

Pour ces variables, on rejette l'hypothèse d'indépendance ( $H_0$ ) : elles présentent une association statistiquement significative avec le fait d'être ou non diagnostiqué gonorrhée.

Variables jugées indépendantes du diagnostic ( $p \geq 0,05$ ) : MTS\_ANT, MTS\_ANT\_GROUP, RAISON, HISTOIRE.

Ici, on ne rejette pas l'hypothèse d'indépendance : on ne dispose pas de preuve statistique suffisante pour conclure qu'elles influencent directement le diagnostic.

D'un point de vue pratique, les variables ayant un lien avec le diagnostic (par exemple, orientation sexuelle, nombre de partenaires, sexe, âge, etc.) méritent une attention particulière : elles pourraient constituer des facteurs de risque ou de protection. Inversement, l'absence d'association statistique pour MTS\_ANT, RAISON ou HISTOIRE de contact peut paraître contre-intuitive (par exemple, on aurait pu s'attendre à une influence de l'historique de MTS).

Modélisation par la régression logistique Comme mentionné dès le début de notre étude, nous cherchons à prédire si un individu sera diagnostiqué gonorrhée ou non, tout en caractérisant le groupe à risque auquel il appartient. Notre variable dépendante est donc la variable DIAGN (diagnostique), et nous retenons toutes les autres variables comme variables explicatives. Avant de commencer la modélisation, nous allons calculer la fréquence de chaque classe dans la variable DIAGN afin de voir l'état d'équilibre entre les deux classes.

TABLE 6 – Répartition des classes dans l'ensemble initial

DIAGN_1.0	Effectif	Pourcentage (%)
0	2368	75.606641
1	764	24.393359

On constate que l'ensemble de données initial présente un déséquilibre de classes : environ 75,6% des individus sont diagnostiqués «0» (pas de gonorrhée), contre seulement 24,4% diagnostiqués «1» (gonorrhée). Ce déséquilibre indique que la classe «1» est moins représentée, ce qui peut avoir un impact sur la performance de la prédiction du modèle (par exemple, l'algorithme pourrait avoir tendance à prédire la classe majoritaire). Pour cela on doit intégrer le poids dans notre modèle de prédiction pour améliorer la capacité à détecter la classe minoritaire.

Par la suite, nous avons calculé le VIF (Variance Inflation Factor) afin de détecter une éventuelle multicollinéarité entre les variables, c'est-à-dire évaluer si l'une d'entre elles peut être prédite à partir des autres.

TABLE 7 – Valeurs du VIF (Variance Inflation Factor)

Variable	VIF
AGE_GROUP_>=30	1.556873
MTS_ANT_GROUP_PAS_ENCORE	1.866298
PARTNER_GROUP_TRÈS_ACTIFS	1.214462
SEXE_2	1.891490
ETAT_C_2.0	1.234815
ETAT_C_3.0	1.184231
ETAT_C_4.0	1.013865
ORIENT_SEX_2.0	2.420621
ORIENT_SEX_3.0	1.120403
RAISON_2.0	<i>inf</i>
RAISON_3.0	1.479381
RAISON_4.0	1.028140
RAISON_5.0	1.022525
HISTOIRE_1.0	<i>inf</i>

On constate que les variables RAISON\_2 et HISTOIRE\_1.0 présentent un VIF infini, indiquant une forte colinéarité. Pour la modélisation, il est donc nécessaire de supprimer l'une de ces deux variables afin d'éviter les problèmes liés à la multicollinéarité.

Après avoir supprimé la variable RAISON\_2, nous avons recalculé le VIF pour l'ensemble des variables restantes. Toutes les valeurs obtenues sont inférieures à 3, ce qui suggère qu'aucun problème majeur de multicollinéarité ne subsiste.

Par la suite, nous avons séparé notre jeu de données en trois ensembles : un ensemble d'entraînement(60%), un ensemble de validation(20%) et un ensemble de test(20%), pour la modélisation et la prédiction. Pour la modélisation, nous avons eu recours à la validation croisée afin de sélectionner les dix variables les plus pertinentes pour la prédiction de la gonorrhée.

# Résultats et interprétations

## 0.5 Description des groupes à risque identifiés

TABLE 8 – Logit Regression Results

Dep. Variable:	DIAGN\1.0	No. Observations:	3132		
Model:	Logit	Df Residuals:	3118		
Method:	MLE	Df Model:	13		
Date:	Tue, 08 Apr 2025	Pseudo R-squ.:	0.05358		
Time:	16:25:20	Log-Likelihood:	-1646.8		
converged:	True	LL-Null:	-1740.1		
Covariance Type:	nonrobust	LLR p-value:	8.053e-33		
\midrule					
=====					
	coef	std err	z	P> z	[0.025
-----					
const	-0.7695	0.096	-8.049	0.000	-0.957
AGE_GROUP_>=30	-0.3999	0.096	-4.163	0.000	-0.588
MTS_ANT_GROUP_PAS ENCORE	0.0744	0.090	0.831	0.406	-0.101
PARTNER_GROUP_TRÈS ACTIFS	0.5455	0.102	5.333	0.000	0.345
SEXE_2	-0.4571	0.142	-3.212	0.001	-0.736
ETAT_C_2.0	-0.2492	0.215	-1.161	0.246	-0.670
ETAT_C_3.0	0.2787	0.216	1.289	0.197	-0.145
ETAT_C_4.0	0.2766	1.089	0.254	0.800	-1.858
ORIENT_SEX_2.0	-0.7045	0.118	-5.990	0.000	-0.935
ORIENT_SEX_3.0	-0.6934	0.172	-4.022	0.000	-1.031
RAISON_3.0	-0.0131	0.100	-0.131	0.896	-0.208
RAISON_4.0	-0.4133	0.346	-1.196	0.232	-1.091
RAISON_5.0	-0.6548	0.460	-1.424	0.154	-1.556
HISTOIRE_1.0	0.2028	0.120	1.689	0.091	-0.033

TABLE 9 – Coefficient, Odds Ratio, Intervalle de Confiance, p-value et significativité

Variable	Coefficient	Odds Ratio	IC Lower	IC Upper
AGE_GROUP_>=30	-0.40	0.67	0.56	0.81
MTS_ANT_GROUP_PAS ENCORE	0.07	1.08	0.90	1.28
PARTNER_GROUP_TRÈS ACTIFS	0.55	1.73	1.41	2.11
SEXE_2	-0.46	0.63	0.48	0.84
ETAT_C_2.0	-0.25	0.78	0.51	1.19
ETAT_C_3.0	0.28	1.32	0.86	2.02
ETAT_C_4.0	0.28	1.32	0.16	11.15
ORIENT_SEX_2.0	-0.70	0.49	0.39	0.62
ORIENT_SEX_3.0	-0.69	0.50	0.36	0.70
RAISON_3.0	-0.01	0.99	0.81	1.20
RAISON_4.0	-0.41	0.66	0.34	1.30
RAISON_5.0	-0.65	0.52	0.21	1.28
HISTOIRE_1.0	0.20	1.22	0.97	1.55

## Interprétation des Odds Ratios pour les avriables significatives

Les variables significatives de la régression logistique influencent la probabilité d'être diagnostiqué positif de la manière suivante :

## Interprétation des Odds Ratios

### — AGE\_GROUP\_>=30 (Odds Ratio = 0.67)

Les individus appartenant au groupe d'âge  $\geq 30$  ans ont environ 33% de chances de moins d'être diagnostiqués positifs par rapport aux individus de moins de 30 ans. Cela indique un effet protecteur associé à l'appartenance au groupe  $\geq 30$  ans.

- **PARTNER\_GROUP\_TRÈS ACTIFS (Odds Ratio = 1.73)**  
Les personnes présentant un nombre élevé de partenaires ont 1.73 fois plus de chances d'être diagnostiquées positives par rapport aux personnes avec un nombre faible de partenaires, ce qui souligne l'importance du nombre élevé de partenaires comme facteur de risque.
- **SEXE\_2(femme) (Odds Ratio = 0.63)**  
Cette variable montre qu'appartenir à la catégorie indiquée par SEXE\_2 est associé à une réduction d'environ 37% des chances d'obtenir un diagnostic positif, indiquant un effet protecteur.
- **ORIENT\_SEX\_2.0(hétérosexuel) (Odds Ratio = 0.49)**  
Les individus de la catégorie ORIENT\_SEX\_2.0 présentent environ 51% de chances en moins d'être diagnostiqués positifs par rapport à la référence(homosexuel), ce qui traduit un fort effet protecteur.
- **ORIENT\_SEX\_3.0(autres) (Odds Ratio = 0.50)**  
De même, un odds ratio de 0.50 signifie que ces individus ont environ 50% de chances de moins d'obtenir un diagnostic positif par rapport à la catégorie de référence(homosexuel), confirmant ainsi un effet protecteur.
- **Variables non significatives**  
Les variables telles que MTS\_ANT\_GROUP\_PAS ENCORE, ETAT\_C\_2.0, ETAT\_C\_3.0, ETAT\_C\_4.0, RAISON\_3.0, RAISON\_4.0, RAISON\_5.0 et HISTOIRE\_1.0 ne présentent pas d'effet statistiquement significatif (p-values supérieures au seuil de 0,05). Cela signifie que nous ne pouvons pas conclure qu'elles influencent de manière certaine le diagnostic positif dans ce modèle.

**Résumé :** Les résultats indiquent que l'appartenance au groupe d'âge  $\geq 30$  ans est protectrice, tandis qu'un nombre élevé de partenaires augmente significativement le risque de diagnostic positif. De plus, les catégories d'orientation sexuelle (ORIENT\_SEX\_2.0 et ORIENT\_SEX\_3.0) ainsi que la variable SEXE\_2 présentent des effets protecteurs.

Ces résultats indiquent que les facteurs tels qu'un nombre élevé de partenaires, l'homosexualité, les jeunes de moins de 30 ans, et les hommes augmentent significativement le risque de diagnostic positif pour la gonorrhée. En revanche, l'absence d'antécédents de MTS agit comme un facteur protecteur.

## Conclusion

En synthèse, ce rapport montre l'importance de cibler des populations particulières pour un dépistage efficace de la gonorrhée. Les facteurs de risque tels que l'âge, l'orientation sexuelle, l'historique de contact avec une personne infectée et le nombre de partenaires ont été mis en évidence comme déterminants majeurs.

La prise en considération de ces données permet de développer des campagnes de prévention et de sensibilisation plus adaptées, contribuant ainsi à la lutte contre la propagation des MTS. Les résultats obtenus ouvrent la voie à des pistes d'amélioration pour les futurs programmes de dépistage, en insistant sur l'importance de la formation des médecins et l'accessibilité des services de santé.