

# Impact de l'information (Q1 avant vs Q2 après)

## Table des matières

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Objectif . . . . .	2
1.2	Méthodes . . . . .	3
<b>2</b>	<b>Données et préparation</b>	<b>5</b>
2.1	Données . . . . .	5
2.2	Préparation des scores . . . . .	5
<b>3</b>	<b>Description</b>	<b>5</b>
3.1	Langue des questionnaires (français vs créole) . . . . .	6
3.2	Appariement . . . . .	6
3.3	Proportion d'appariement (Q1 $\rightarrow$ Q2) . . . . .	8
3.4	Schéma de flux (effectifs) . . . . .	10
3.5	Attrition selon l'âge (Q1 appariées vs Q1 non appariées) . . . . .	11
<b>4</b>	<b>Résultats principaux</b>	<b>13</b>
4.1	Résultats clés . . . . .	13
4.2	Appariement et population analysée . . . . .	13
4.3	Distribution des scores initiaux (Q1) . . . . .	15
4.3.1	Boxplots côte à côte (score global) . . . . .	16
4.3.2	Visualisation des trajectoires individuelles . . . . .	17
4.3.3	Diagramme spaghetti (évolution individuelle) . . . . .	17
4.3.4	Diagramme spaghetti coloré (par direction du gain) . . . . .	19
4.3.5	Nuage de points Q1 vs Q2 (appariés) . . . . .	20
4.3.6	Waterfall plot des gains (Q2 $-$ Q1) . . . . .	21
4.3.7	Gain en fonction du score initial (effet plafond) . . . . .	22
4.3.8	Analyse et interprétation (avant/après) . . . . .	23
4.3.9	Test statistique (test t apparié / Student) . . . . .	24
4.3.10	Tableau de synthèse (analyse principale) . . . . .	24
4.3.11	Le gain est-il cliniquement pertinent ? . . . . .	25

4.3.12	Taille d'effet : d de Cohen . . . . .	26
4.3.13	Distribution des gains (Q2 – Q1) . . . . .	27
4.3.14	Proportion améliorée / stable / diminuée . . . . .	28
4.3.15	Résumé des gains . . . . .	29
4.3.16	Pertinence clinique : proportion avec gain $\geq 3$ points . . . . .	29
4.3.17	Probabilité d'un gain $\geq 3$ points selon le score initial . . . . .	29
4.3.18	Test de signe (proportion de paires améliorées) . . . . .	32
4.4	Score initial (Q1) selon l'âge . . . . .	33
4.5	Exploration : gain global selon l'âge . . . . .	33
4.5.1	Répondeurs (gain $\geq 3$ ) selon l'âge (exploratoire) . . . . .	34
4.6	Scores initiaux par question (Q1) . . . . .	36
<b>5</b>	<b>Analyse par question</b>	<b>42</b>
5.1	Résultats par question . . . . .	43
5.2	Synthèse des gains par question . . . . .	44
5.3	Tests par question (appariés) . . . . .	45
5.4	Graphique : score moyen par question (avant vs après) . . . . .	46
5.4.1	Version enrichie (avec valeur du gain) . . . . .	47
5.5	Autre visualisation de l'amélioration . . . . .	48
5.6	Scores et gains par question (1 graphique par question) . . . . .	49
5.7	Vue d'ensemble : distribution Q1 vs Q2 . . . . .	55
5.8	Forest plot récapitulatif (gain global + par question) . . . . .	55
<b>6</b>	<b>Biais de sélection potentiel : Q1 appariées vs Q1 non appariées</b>	<b>56</b>
<b>7</b>	<b>Discussion</b>	<b>59</b>
7.1	Résultats principaux . . . . .	59
7.2	Limites méthodologiques . . . . .	59
7.3	Recommandations pour une future étude . . . . .	60
7.4	Version thèse (paragraphe) . . . . .	60
<b>8</b>	<b>Annexes techniques (détails statistiques)</b>	<b>61</b>
8.1	Diagnostic visuel (gain global) . . . . .	61
8.2	Bootstrap . . . . .	63
8.2.1	Densités (avant bootstrap) et densités des moyennes (bootstrap) . . . . .	63

# 1 Introduction

## 1.1 Objectif

Objectif principal : évaluer l'impact d'une information sur les connaissances des risques de l'alcool pendant la grossesse, en comparant le **score global** avant (Q1) et après (Q2) information,

chez les mêmes participantes.

Objectifs secondaires : décrire (i) la proportion de participantes ayant un gain  $\geq 3$  points, (ii) l'évolution individuelle (spaghetti), (iii) les évolutions **par question**, et (iv) l'hétérogénéité des scores/gains selon l'âge (exploratoire).

## 1.2 Méthodes

### **i** Approche méthodologique en résumé

Il s'agit d'une étude **avant–après** visant à mesurer l'évolution des connaissances des participantes suite à une intervention d'information. La même population a répondu à un questionnaire avant (Q1) puis après (Q2).

L'analyse repose sur la comparaison des scores Q1 et Q2 **chez des patientes appairées**, ce qui est la méthode la plus fiable pour mesurer un changement.

En l'absence de critère d'identification unique, l'appariement a été réalisé via des identifiants partiels (âge et commune de résidence), puis validé manuellement. Seules les paires complètes (Q1 et Q2) ont été retenues pour l'analyse principale.

- **Tests statistiques** : On utilise le **test t apparié** pour vérifier que l'amélioration n'est pas due au hasard. *Une méthode de bootstrap permet de réaliser un test t sans présupposer de normalité des données.*
- **Quantification de l'effet** : On calcule le **gain moyen** et son **intervalle de confiance à 95%** pour connaître la taille de l'amélioration et sa plage de valeurs plausibles. Le **bootstrap** valide ces résultats sans présupposer de la normalité des données.
- **Pertinence et effet plafond** : On analyse la proportion de participantes avec un **gain  $\geq 3$  points** et on utilise une **régression logistique** pour comprendre comment la probabilité d'un "gros gain" diminue quand le score de départ est déjà élevé.
- **Analyses exploratoires** : On regarde ensuite les différences par **âge** et par **question** pour affiner la compréhension.

### Texte complet :

Il s'agit d'une étude avant–après à données appariées évaluant l'impact d'une information sur les connaissances des participantes sur les risques de l'alcool pendant la grossesse, mesurées par deux passations d'un questionnaire (Q1 avant, Q2 après).

En l'absence d'information complète sur l'identité des patientes, une procédure d'appariement a été mise en place via des identifiants partiels (âge et commune de résidence), puis validé manuellement. Seules les paires complètes ont été retenues pour l'analyse principale. Le critère de jugement principal est le score global, calculé à partir des réponses à l'aide d'une clé de

correction et d'un barème prédéfinis : les points des 8 questions sont sommés (total /40), puis divisés par 2 pour obtenir une note sur 20. Pour les questions à réponses multiples, un point est attribué à chaque proposition correcte selon le barème ; aucune pénalité n'est appliquée en cas de réponses supplémentaires non attendues. Les scores par question (en points, de 0 au maximum de la question) et le gain individuel ( $Q2 - Q1$ ) ont également été calculés ; pour comparer des questions de barèmes différents, les gains ont aussi été exprimés en pourcentage du barème.

Les variables quantitatives sont décrites par moyenne  $\pm$  écart-type (et médiane [IQR] lorsque pertinent) et les variables qualitatives par effectif et pourcentage. L'effet global de l'intervention sur le score global a été évalué en comparant Q1 et Q2 au sein des mêmes participantes, à l'aide d'un test t apparié (Student), dans le sens "après > avant", en appliquant une procédure de *bootstrap* (1000 réplifications) pour s'affranchir de l'hypothèse de normalité des données. La pertinence clinique a été décrite en (i) comparant le gain moyen au seuil de 3 points et (ii) calculant la proportion de participantes atteignant un gain  $\geq 3$  points (avec intervalle de confiance binomial). Un test de signe (binomial) a également été utilisé pour comparer la fréquence des améliorations (gain  $> 0$ ) à celle des diminutions (gain  $< 0$ ), en ignorant les égalités. Enfin, une analyse exploratoire a estimé la probabilité d'un gain  $\geq 3$  points en fonction du score initial, de l'âge (approché en années) et de la langue du questionnaire (français vs créole), via une régression logistique multivariable.

Des analyses exploratoires ont étudié l'hétérogénéité des scores initiaux et des gains selon la tranche d'âge (tests de Kruskal–Wallis et modèle linéaire ajusté sur le score initial). Enfin, des analyses par question ont comparé les scores avant/après (tests appariés) et décrit les distributions de scores en points (proportions avec intervalles de confiance de Wilson), avec des tests de McNemar par niveau de score ( $p < 0,05$ ). Ces analyses "par question" sont présentées à titre exploratoire, sans correction de multiplicité. Toutes les analyses ont été conduites sous R 4.5.2, avec un seuil de significativité de 5% ( $\alpha = 0,05$ ) et des intervalles de confiance à 95%.

## 2 Données et préparation

### 2.1 Données

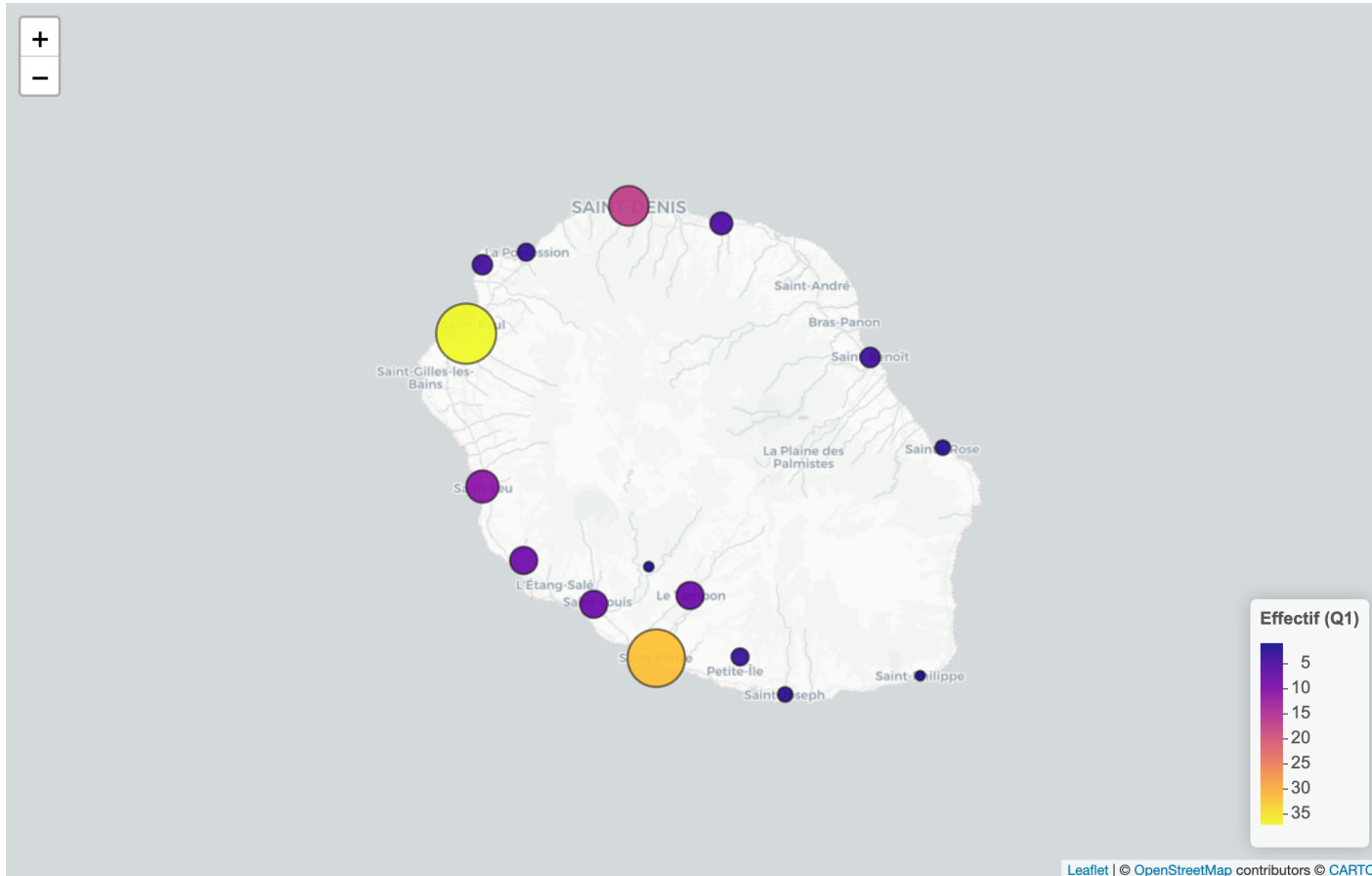
Ce rapport utilise le fichier apparié `Q1_Q2_apparie.csv` (1 ligne = 1 participante du Q1, appariée à une ligne du Q2 quand possible).

### 2.2 Préparation des scores

## 3 Description

La carte suivante montre la répartition des participantes sur l'île de La Réunion.

Chaque commune est représentée par un marqueur, dont la couleur et la taille reflètent le nombre de répondantes.



### 3.1 Langue des questionnaires (français vs créole)

Les questionnaires étaient disponibles en français et en créole réunionnais.

#### Langue du questionnaire (Q1)

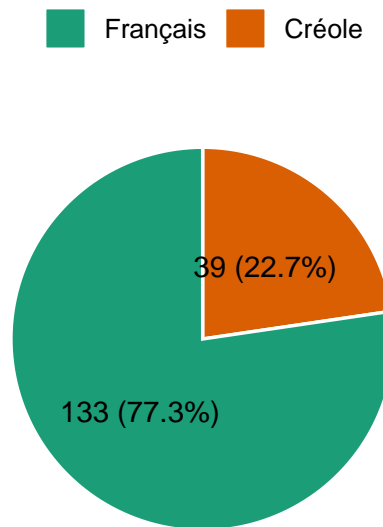


Figure 1: Langue du questionnaire (Q1)

### 3.2 Appariement

Récapitulatif :

- Q1 :  $n = 172$ ,
- Q2 :  $n = 106$ ,
- **paires appariées**  $n = 103$  (soit 59.9% des Q1).

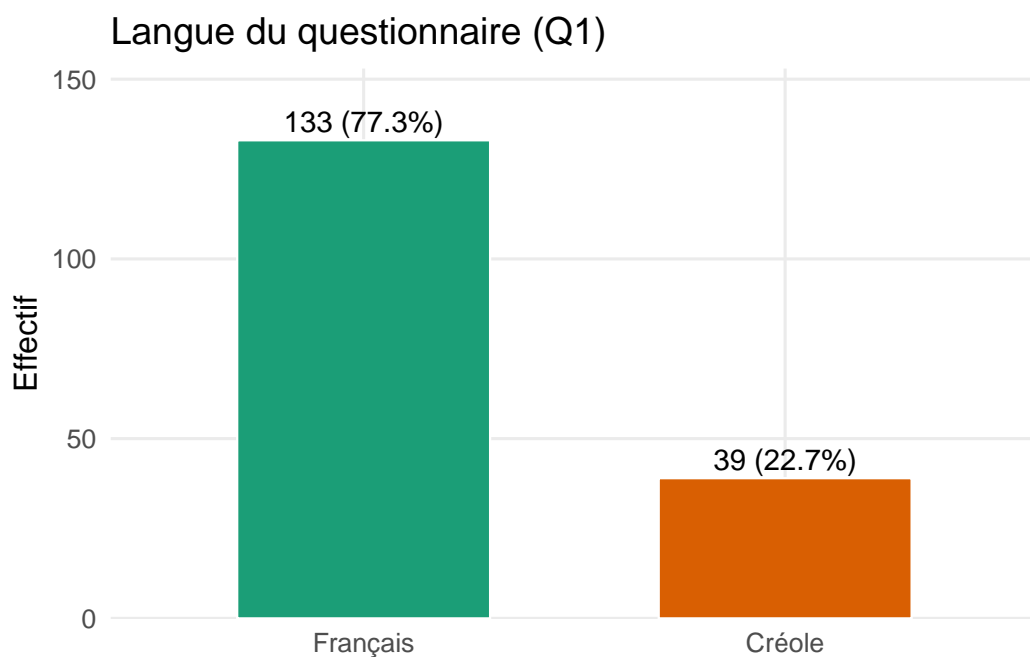


Figure 2: Langue du questionnaire Q1 (barplot)

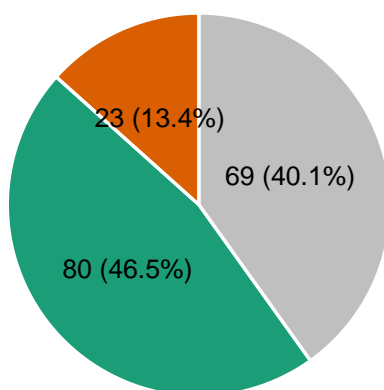
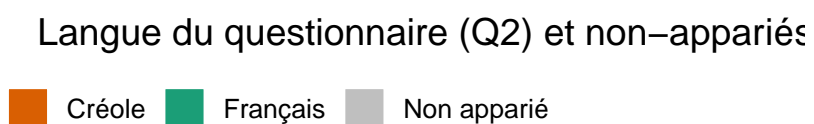


Figure 3: Langue du questionnaire (Q2) et attrition

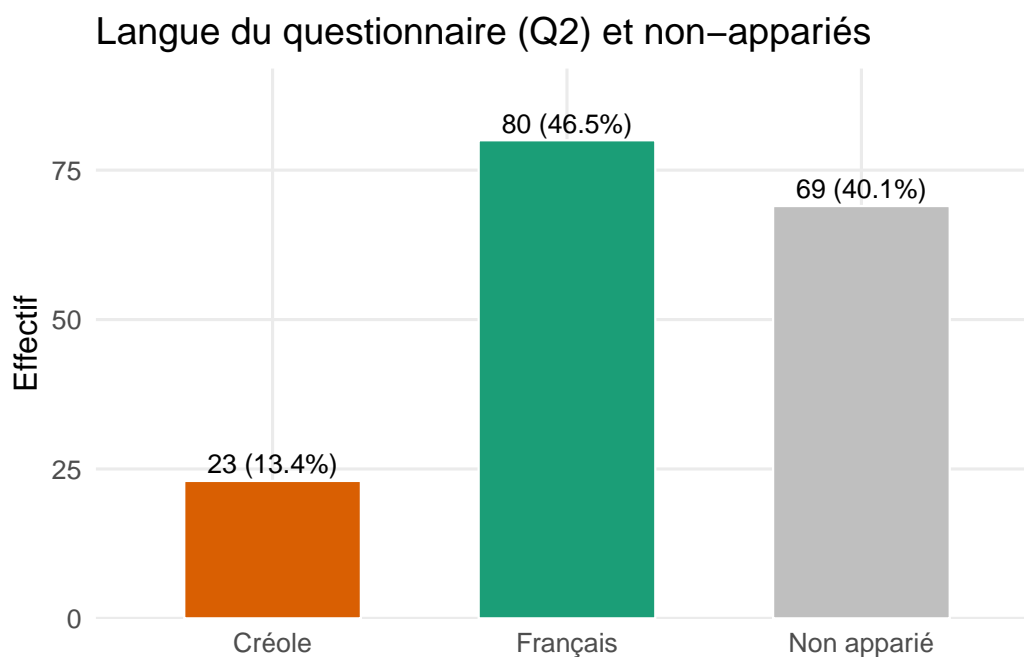
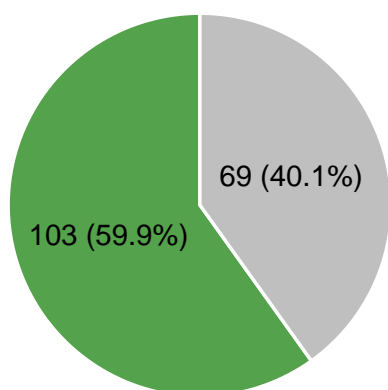


Figure 4: Langue du questionnaire Q2 et non-appariés (barplot)

### 3.3 Proportion d'appariement (Q1 → Q2)

#### Proportion d'appariement

■ Appariées ■ Non appariées



Ici, **103/172** réponses du Q1 sont appariées (soit 59.9%), et l'on retrouve **103/106** réponses du Q2 (soit 97.2%).



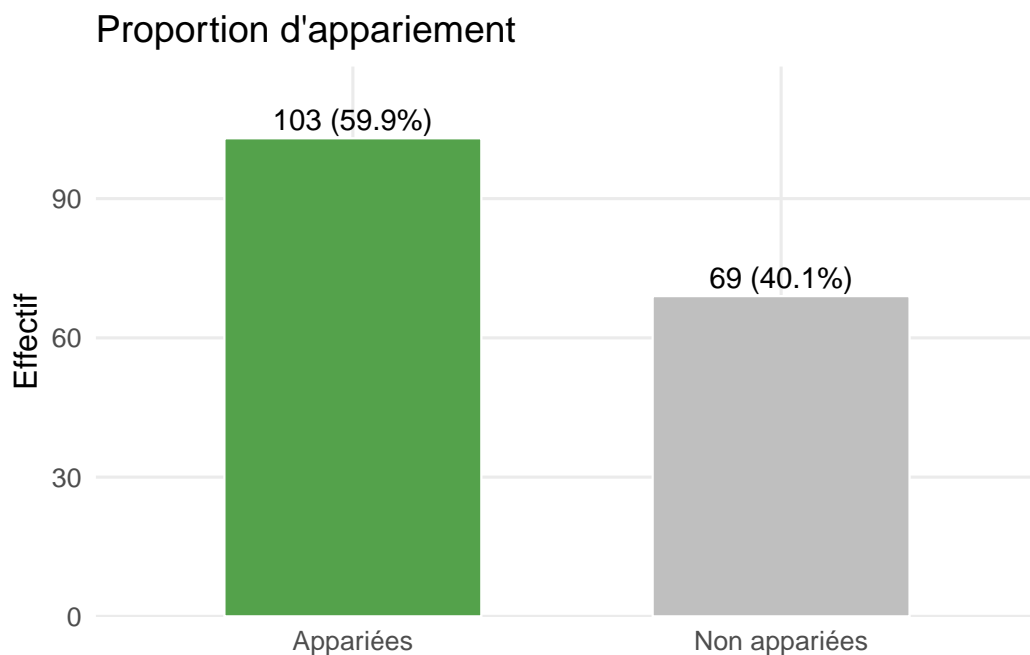


Figure 5: Proportion d'appariement (barplot)

(Donc 103 des 106 réponses du Q2 ont pu être appariées, soit un taux d'appariement de 97.2% du côté Q2.)

Les patientes n'ayant pas pu être appariées n'ont pas renseignées de commune ou d'âge compatible, ou ont des informations incohérentes avec les données du Q1.

L'analyse principale avant-après porte sur l'échantillon "Appariées".

### 3.4 Schéma de flux (effectifs)

Le schéma ci-dessous résume les effectifs à chaque étape (Q1 total → paires analysées → non appariées).

#### Flux des participants

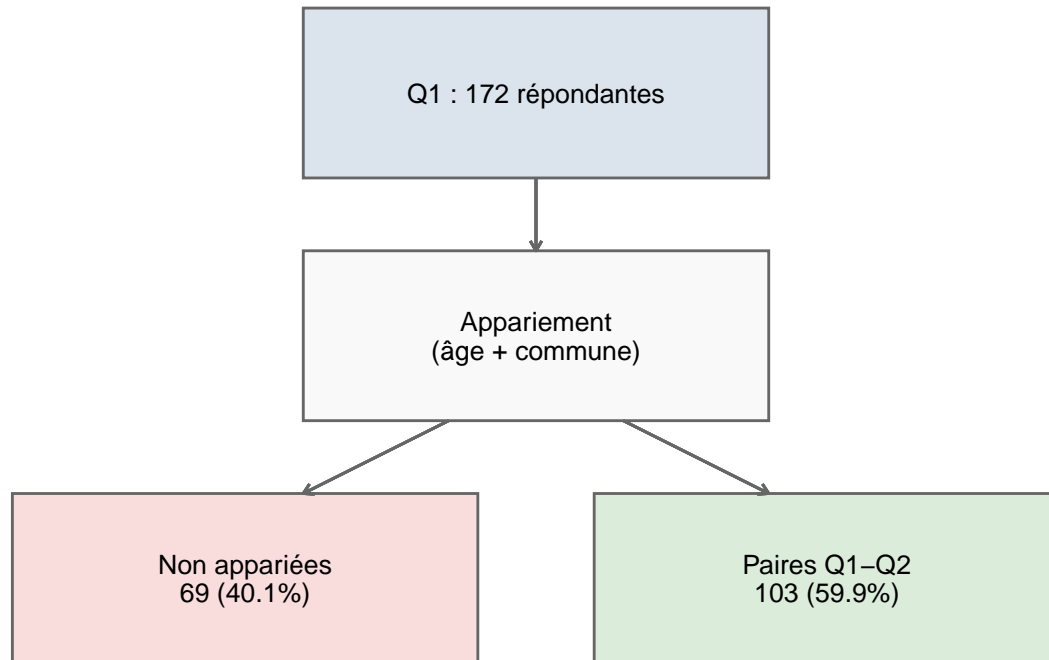


Figure 6: Diagramme de flux simplifié

Interprétation :

Même interprétaion que le camembert précédent.

Des patientes qui ont répondu au 2ème questionnaire, 103/106 réponses ont été appariées (soit 97.2%).

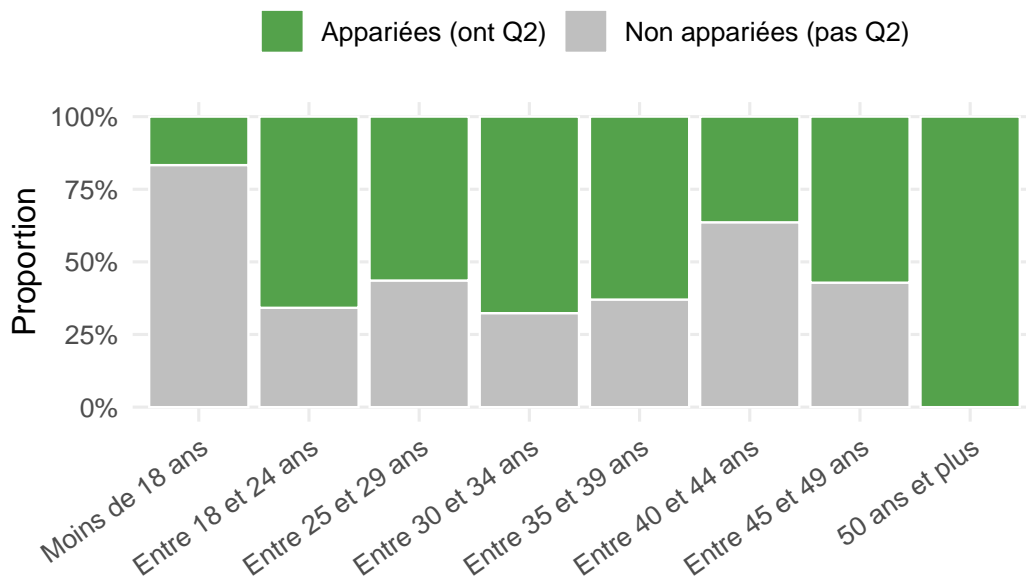
### 3.5 Attrition selon l'âge (Q1 appariées vs Q1 non appariées)

= Y a-t-il plus de pertes au suivi (Q2 manquant) dans certaines tranches d'âge ?

- Via un test du chi2 d'indépendance entre la tranche d'âge et le statut d'appariement (Q1 avec ou sans Q2) : p-value > 0.9 donc pas d'impact évident de l'âge sur l'attrition.
- Visuellement : pas d'impact de l'âge sur l'attrition non plus.

Graphique pas obligatoire : en gros y a pas de différence des “perdus de vue” selon l'âge.

#### Attrition selon l'âge : proportion de Q1 avec ou sans Q2



Même graphique mais avec les effectifs en ordonnées :

## Attrition selon l'âge : effectifs de Q1 avec ou sans Q2

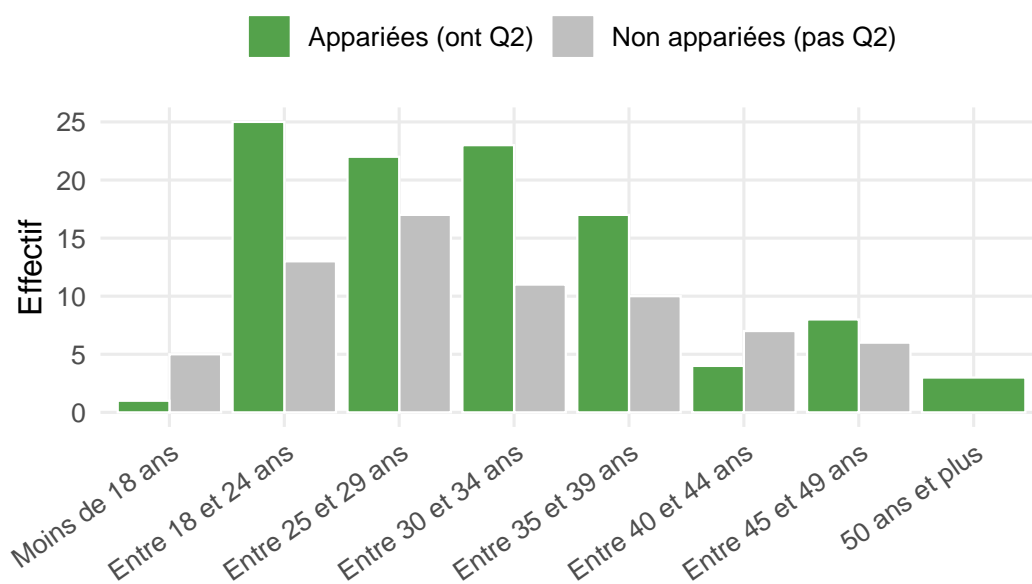


Table 1

## Synthèse des résultats clés

Sur **103** participantes appariées

n (paires)	Moyenne Q1 (/20)	Moyenne Q2 (/20)	Gain moyen (Q2–Q1)	Proportion gain $\geq 3$
103	15.257	18.223	2.966	48.544%

## 4 Résultats principaux

### 4.1 Résultats clés

Le gain moyen entre les groupes Q1 et Q2 (appariées) est à 2,97 points avec un IC à [2,13 ; 3,81].

La proportion de participantes avec un gain  $\geq 3$  points est de 48.5%.

### 4.2 Appariement et population analysée

Principaux résultats en langage simple (échantillon apparié, n = 103)

1. **L’information a-t-elle amélioré les connaissances ?**  
**Oui.**

En moyenne, les participantes ont gagné **2,97 points** sur 20 après information. Cette amélioration est **statistiquement significatif** (t-test apparié unilatéral :  $p = < 1e-4$ ), ce qui veut dire qu’il est très improbable qu’elle soit due au hasard.

2. **L’amélioration est-elle “cliniquement pertinente” (gain d’au moins 3 points) ?**

**En moyenne : presque.**

Le gain moyen est de **2,97** points, donc **très proche** du seuil de 3 points.

L’IC95% bilatéral du gain moyen est [**2,13 ; 3,81**], ce qui est compatible avec un gain moyen autour de 3. En revanche, **48.5%** des participantes ont atteint ou dépassé **3 points** de gain.

3. **Pourquoi le gain moyen n’est-il pas plus élevé ?**

Principalement à cause d’un **effet plafond** : le score de départ était déjà élevé (moyenne Q1 = **15,26/20**).

Quand on part déjà haut, il est mathématiquement plus difficile de gagner beaucoup de points.

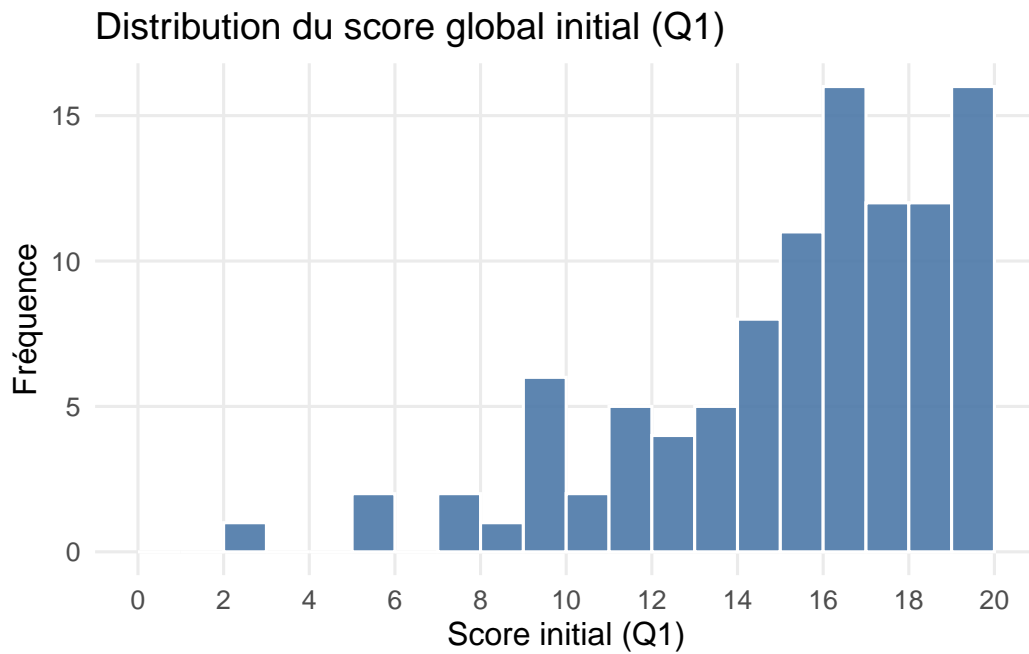
4. **Y a-t-il des différences selon l'âge ?**

**Non, pas de manière évidente.**

La comparaison des gains selon les tranches d'âge est non statistiquement significatif (Kruskal-Wallis :  $p = 0.3518$ ).

**Conclusion simple :** l'information améliore les connaissances en moyenne ; le gain moyen est très proche de 3 points, et le niveau initial déjà élevé limite la marge de progression.

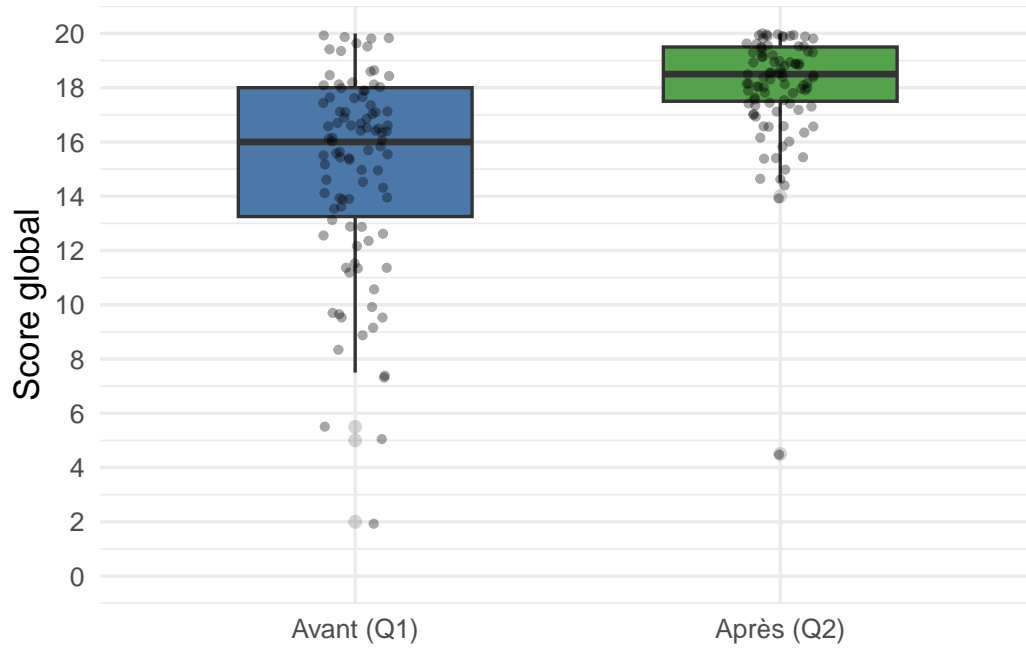
### 4.3 Distribution des scores initiaux (Q1)



Interprétation :

Cet histogramme décrit le niveau de connaissances initial global et permet d'apprécier la dispersion, ainsi qu'un éventuel effet plafond si de nombreux scores sont proches de 20.

#### 4.3.1 Boxplots côte à côte (score global)



Interprétation (boxplots) :

On observe un déplacement global vers des scores plus élevés après l'information. La médiane passe de 16 (Q1) à 18.5 (Q2), et l'intervalle interquartile se déplace dans le même sens.

Les scores aux Q2 se “resserrent” : ils tendent à être plus concentrés autour de la médiane, avec moins de valeurs basses.



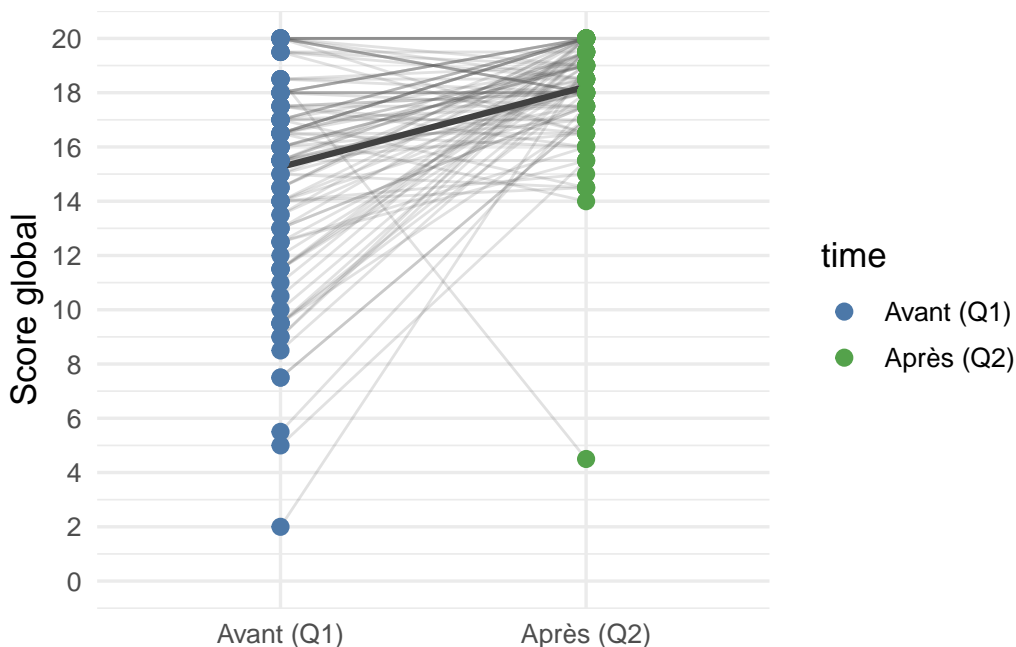
### 4.3.2 Visualisation des trajectoires individuelles

Les graphiques suivants (diagrammes “spaghetti”, nuages de points, ou “waterfall”) ont tous pour but de dépasser l’analyse de la moyenne pour explorer la **dynamique des changements individuels**.

Ils permettent de répondre à des questions comme : “La majorité des participantes a-t-elle progressé ?”, “Y a-t-il des personnes pour qui le score a baissé ?”, “Les gains sont-ils homogènes ou très variables ?”.

Chaque graphique montre globalement la même information, mais avec des mises en forme différentes.

### 4.3.3 Diagramme spaghetti (évolution individuelle)



Interprétation (spaghetti) :

La majorité des trajectoires individuelles monte (gain > 0 : 70 participantes), avec 19 baisses et 14 stabilités. La tendance moyenne est à l’augmentation entre Q1 et Q2.

Chaque ligne grise fine représente la trajectoire d’une personne, et la ligne noire épaisse la **tendance moyenne** du groupe. On voit que la grande majorité des lignes **montent**, ce qui suggère un effet positif global de l’intervention, qui va maintenant être formalisé par les tests statistiques.

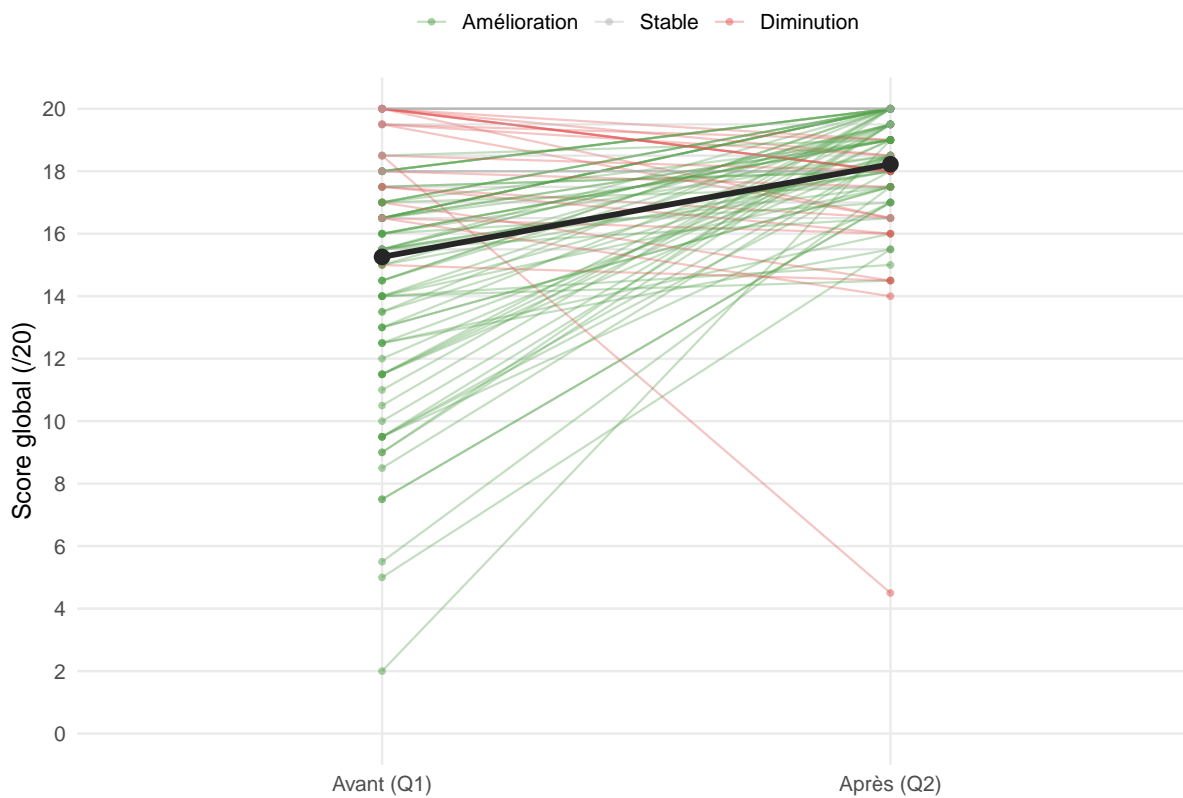
On ne peut pas trop expliquer la ligne qui descend isolée : peut être quelqu'un qui a fermé le questionnaire rapidement sans le remplir sérieusement.

#### 4.3.4 Diagramme spaghetti coloré (par direction du gain)

Ce graphique reprend le diagramme spaghetti ci-dessus, mais chaque trajectoire est colorée selon la direction du changement : vert pour les améliorations, rouge pour les diminutions, gris pour les stabilités.

Trajectoires individuelles colorées par direction du gain

Vert = amélioration (70), gris = stable (14), rouge = diminution (19)

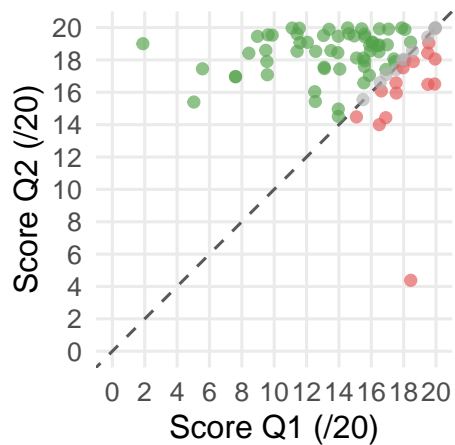


#### 4.3.5 Nuage de points Q1 vs Q2 (appariés)

Scores individuels : Q1 vs Q2

Au-dessus de la diagonale : amélioration ; en c

● Amélioration ● Diminution ● Stable



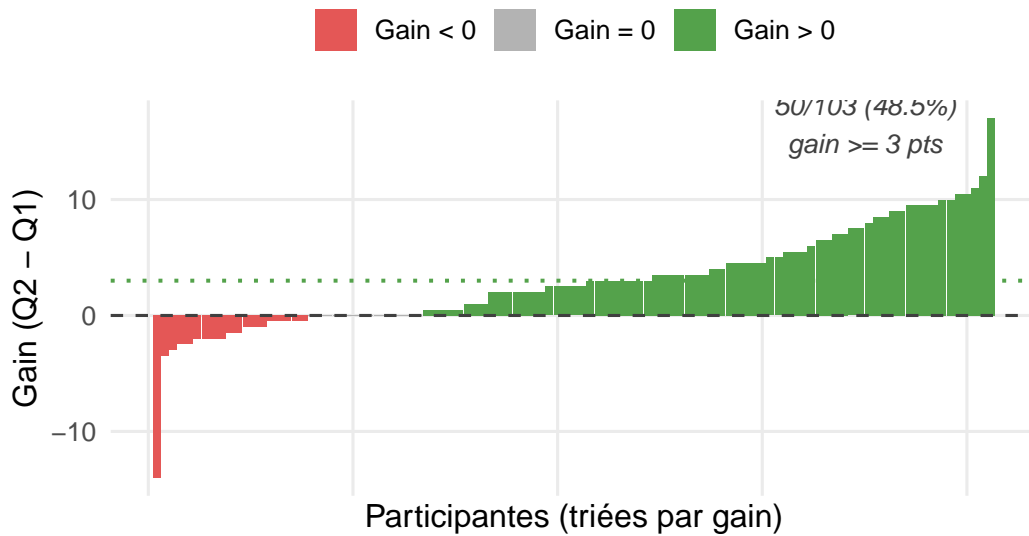
Interprétation :

Ce graphique visualise directement l'évolution individuelle. Si la majorité des points est au-dessus de la diagonale, l'amélioration est majoritaire.

#### 4.3.6 Waterfall plot des gains (Q2 – Q1)

##### Gains individuels triés (waterfall)

Ligne pointillée verte : seuil clinique de 3 points



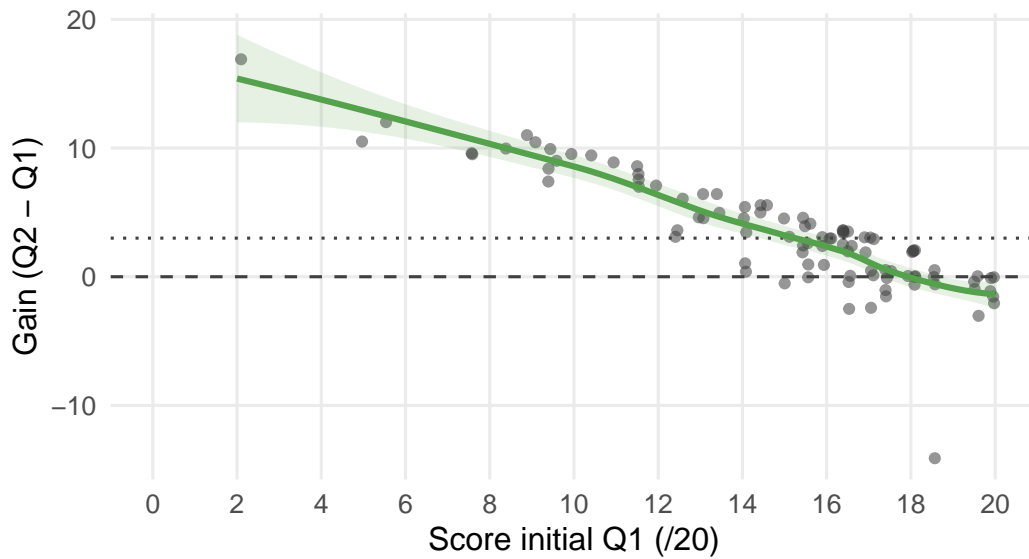
Interprétation :

Ce graphique montre l'ampleur et la variabilité des gains. On visualise aussi la proportion de participants dépassant le seuil de 3 points.

#### 4.3.7 Gain en fonction du score initial (effet plafond)

##### Gain selon le score initial

Tendance attendue : gain plus faible quand le score initial est déjà



Interprétation : La courbe diminue quand le score initial augmente, ce qui illustre un **effet plafond** : plus on part haut, moins on peut progresser.

La plupart des points sont à droite de l'axe des  $x$ , donc concentrés sur les scores élevés. Cela reflète le fait que les participantes avaient déjà un bon niveau de connaissances initial (Q1), limitant la marge de progression.

#### 4.3.8 Analyse et interprétation (avant/après)

Sur les participantes appariées ( $n = 103$ ), le score moyen passe de **15.26** (avant) à **18.22** (après), soit un gain moyen de **2.97** points.

Avec un seuil de pertinence clinique fixé à **3 points**, le gain moyen est **très légèrement en dessous** de ce seuil (gain moyen = 2.97, IC95% bilatéral = [2.13 ; 3.81]).

#### 4.3.9 Test statistique (test t apparié / Student)

Le test principal ci-dessous est un **test t apparié** (Student) qui compare la moyenne des différences à 0 (après - avant). Par défaut, le test est unilatéral (après > avant).

##### Note

À la question : pourquoi avoir choisi ce test :

- Test t car les scores sont continus. Ils ne suivent pas forcément une loi normale, mais on applique ici une procédure de *bootstrap* à 1000 réplifications pour estimer l'IC95% du gain moyen, ce qui rend le test robuste aux écarts à la normalité.
- Apparié car on compare les scores avant/après **chez les mêmes participantes**.
- On aurait pu utiliser un test non paramétrique (Wilcoxon signé-rang), mais (i) le test t est plus puissant si les conditions sont remplies car compare des moyennes et pas des rangs, (ii) l'utilisation d'un test non paramétrique rendrait inconcevable la réalisation future d'une régression logistique multivariable sur le gain, car les régressions sur des données appariées nécessitent des résidus normaux (ce qui n'est pas le cas avec des rangs).

On teste l'hypothèse d'une amélioration des scores après information.

#### 4.3.10 Tableau de synthèse (analyse principale)

Ce tableau résume les résultats principaux sur l'échantillon apparié. Les tests sont rapportés en **bilatéral** (lecture standard) et en **unilatéral** (après > avant, cohérent avec l'hypothèse attendue).

n (paires)	Q1 (moyenne ± SD)	Q2 (moyenne ± SD)	Gain moyen	IC95% gain	p t-test (bilat.)
103	15.26 ± 3.77	18.22 ± 2.05	2.97	[2.13 ; 3.81]	< 1e-4

Interprétation (t-test apparié, données observées) :

Le gain moyen est de 2.97 points.

Le test t apparié est statistiquement significatif ( $p = < 1e-4$ ), en faveur d'une **augmentation** après intervention.

L'IC95% bilatéral du gain moyen est [2.13, 3.81] points.



#### 4.3.11 Le gain est-il cliniquement pertinent ?

On teste si le gain moyen est supérieur au seuil clinique de 3 points.

One Sample t-test

```
data: data_matched$score_diff
t = -0.080315, df = 102, p-value = 0.5319
alternative hypothesis: true mean is greater than 3
95 percent confidence interval:
 2.263719      Inf
sample estimates:
mean of x
 2.966019
```

Interprétation (seuil clinique) :

Le gain moyen observé est de 2.97 points, donc **très légèrement en dessous** du seuil clinique de 3 points.

Le test t unilatéral comparant le gain moyen à 3 points est non statistiquement significatif ( $p = 0.5319$ ).

Donc l'augmentation entre Q1 et Q2 est **pas cliniquement pertinente** selon ce critère.

#### 4.3.12 Taille d'effet : d de Cohen

##### **i** Qu'est-ce que le d de Cohen ?

Le **d de Cohen** est un indicateur de **taille d'effet** qui quantifie l'ampleur d'une différence entre deux mesures, **indépendamment de la taille de l'échantillon**. Contrairement à la p-value (qui dit seulement si l'effet est "probablement réel"), le d de Cohen dit **à quel point l'effet est grand**.

**Calcul :** pour des données appariées,  $d = \frac{\bar{x}_{\text{gain}}}{SD_{\text{gain}}}$ , c'est-à-dire le gain moyen divisé par l'écart-type des gains.

**Interprétation (conventions de Cohen) :**

d	Interprétation
< 0,2	Effet négligeable
0,2 – 0,5	Effet <b>petit</b>
0,5 – 0,8	Effet <b>moyen</b>
> 0,8	Effet <b>grand</b>

Ces seuils sont des repères conventionnels ; l'interprétation doit toujours tenir compte du contexte (ici, un QCM de connaissances avec effet plafond).

d de Cohen	IC95%	Interprétation	Gain moyen	SD du gain
0.691	[0.476 ; 0.906]	Effet moyen	2.97	4.29

Interprétation :

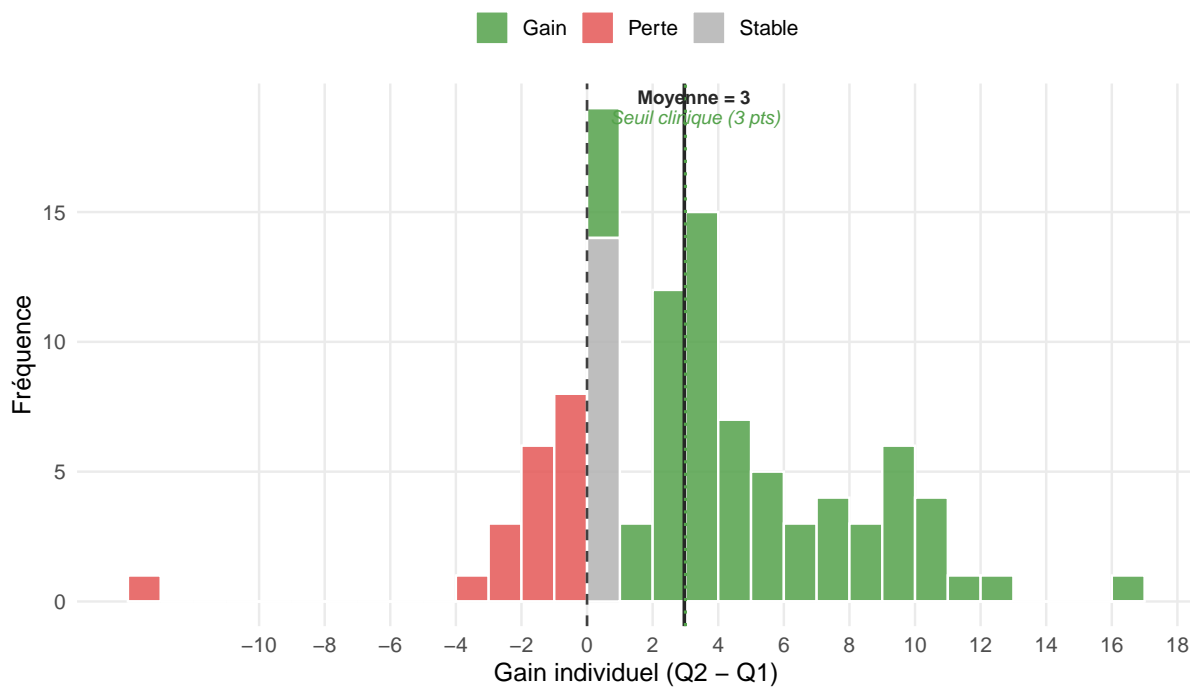
Le d de Cohen est de **0.69** (IC95% : [0.48 ; 0.91]), ce qui correspond à un effet **moyen** selon les conventions de Cohen. Cela signifie que le gain moyen représente environ **0.69 écart-type** de la variabilité interindividuelle des gains.

Dans le contexte d'une intervention éducative sur les connaissances (avec un effet plafond notable), cet effet moyen est à interpréter en tenant compte du fait que les scores initiaux étaient déjà élevés, limitant mécaniquement la marge de progression.

Globalement : il s'agit d'un résultat complémentaire qu'il n'est pas forcément utile d'inclure dans les résultats et dans la thèse, mais s'ils le demandent, c'est là!

#### 4.3.13 Distribution des gains (Q2 – Q1)

Distribution des gains (n = 103 paires)



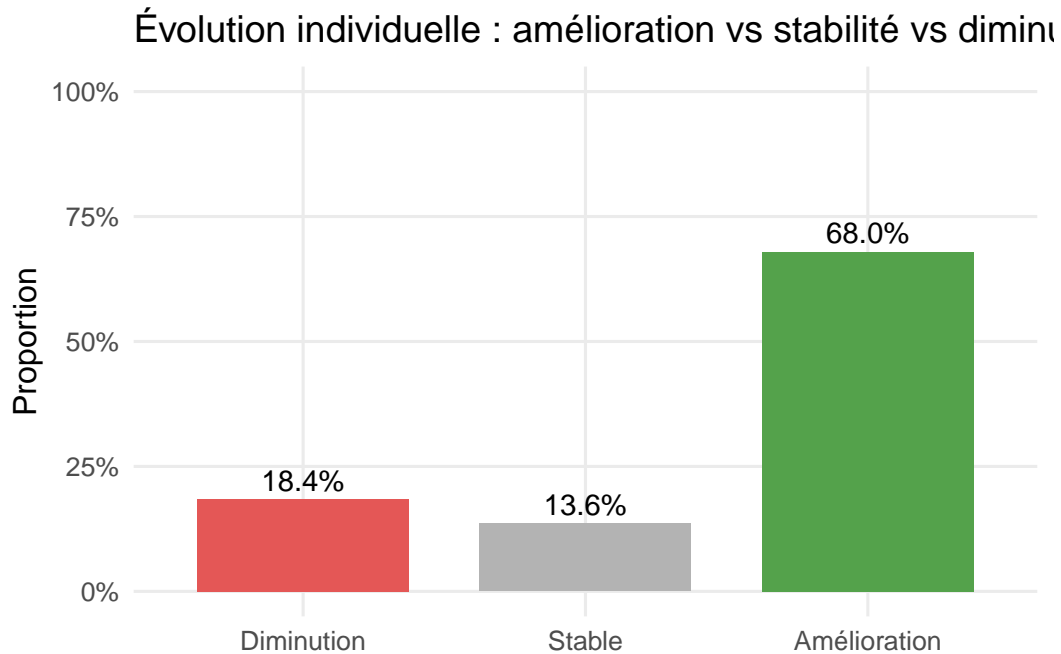
Interprétation :

Cette distribution aide à visualiser l'hétérogénéité des réponses individuelles (certaines participantes progressent beaucoup, d'autres peu, et quelques-unes diminuent). Les barres vertes (gains), grises (stabilité) et rouges (pertes) permettent de voir immédiatement la répartition.

La ligne pleine indique la moyenne, la ligne pointillée verte le seuil clinique de 3 points.

Globalement, on observe une majorité de participantes avec un gain positif !

#### 4.3.14 Proportion améliorée / stable / diminuée



Là aussi, la majorité des patientes s'améliorent (le graphique ici ne tient pas compte de l'ampleur de l'amélioration, par rapport au graphique précédent)

#### 4.3.15 Résumé des gains

#### 4.3.16 Pertinence clinique : proportion avec gain $\geq 3$ points

Interprétation (proportion avec gain  $\geq 3$ ) :

- 50/103 participantes (48.5%) atteignent au moins 3 points.
- IC95% de cette proportion : [0.386, 0.586].

#### 4.3.17 Probabilité d'un gain $\geq 3$ points selon le score initial

##### **i** Objectif de cette analyse

Cette section vise à quantifier l'**effet plafond**.

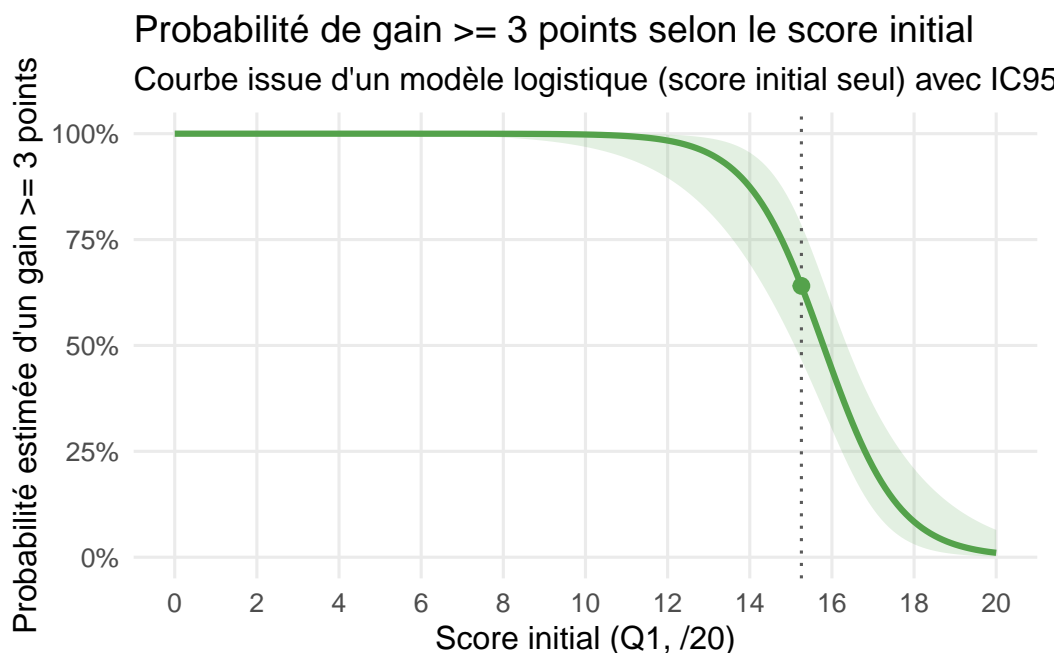
On s'attend à ce qu'il soit plus difficile d'obtenir un gros gain ( $\geq 3$  points) si le score de départ est déjà élevé.

Le modèle de régression logistique estime la **probabilité** que cela arrive, en fonction du score de départ.

Le graphique qui suit est la visualisation de cette probabilité.

Objectif : estimer, à partir des données, la **probabilité** d'obtenir un gain d'au moins 3 points ( $Q2 - Q1 \geq 3$ ) en fonction du score initial égal à la moyenne observée au Q1 dans l'échantillon apparié (15,26/20).

Utilisation d'une régression logistique binaire.



Interprétation :

- Globalement, la courbe est décroissante : quand le score initial augmente, la probabilité d'obtenir un gain  $\geq 3$  diminue (effet plafond).

Pour un score initial égal à la **moyenne du Q1 dans l'échantillon apparié (15,26/20)**, la probabilité estimée d'avoir une amélioration  $> 3$  points est d'environ **64.1%** (IC95% : [46.6%, 78.5%]).

Analyse complémentaire (exploratoire) :

Permet de répondre à la question : **Est-ce que l'âge et/ou la langue du questionnaire influencent la probabilité d'un gain  $\geq 3$  points, à score initial comparable ?**

- on ajuste aussi un modèle de régression logistique multivariable "complet" (score initial + âge + langue) pour vérifier si l'âge et/ou la langue apportent une information supplémentaire sur la probabilité d'un gain  $\geq 3$  points, à score initial comparable.

Variable	OR	IC95% bas	IC95% haut	p
Score initial (Q1) (par +1 point)	0.3352401	0.2131620	0.5272324	$< 1e-4$
Âge (par +1 an)	1.0399252	0.9664946	1.1189347	0.2947
Langue: Créole (vs Français)	0.4732596	0.0618763	3.6197183	0.4711

Interprétation (modèle complet) : dans ces données, on n'observe **pas d'effet clair de l'âge** ni de la **langue du questionnaire** (IC95% des OR recouvrant 1 et  $p > 0,05$ ). En revanche, le

**score initial** est fortement associé à la probabilité d'un gain  $\geq 3$  : plus le score de départ est élevé, plus la probabilité d'un “gros gain” diminue (effet plafond).

Concrètement, l'OR associé au score initial est d'environ **0.34** par +1 point : cela correspond à une baisse d'environ **66%** des “chances” (odds) d'atteindre un gain  $\geq 3$  pour chaque point supplémentaire au score initial.

#### 4.3.18 Test de signe (proportion de paires améliorées)

Le test de signe compare le nombre de paires où le score augmente vs diminue (les égalités sont ignorées).

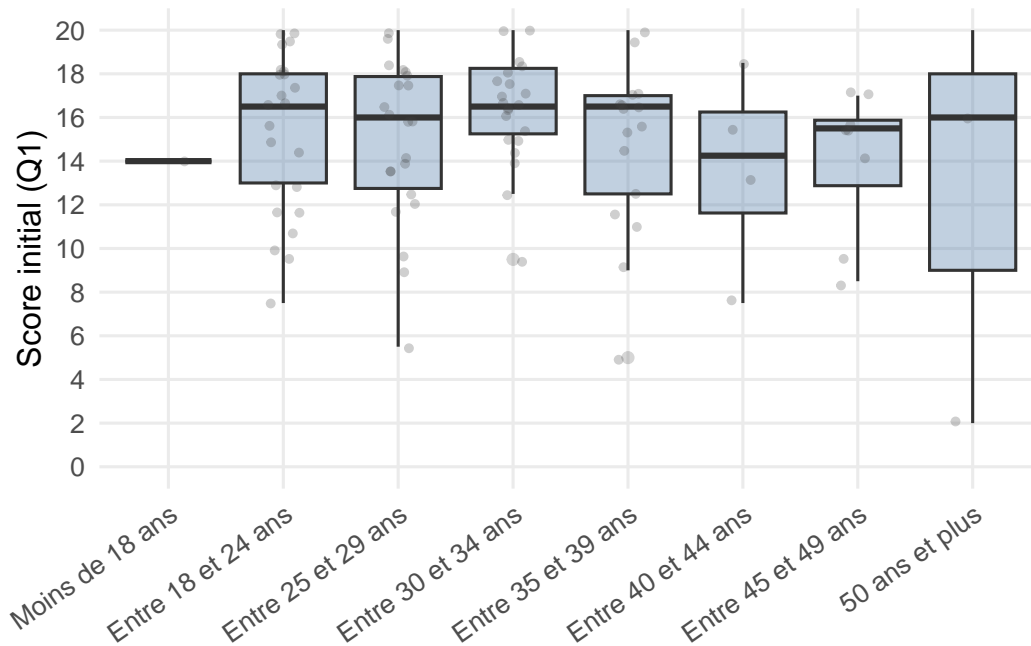
Présentation des résultats en tableau :

Nombre de paires	Paires avec gain > 0	Paires avec gain < 0	Paires avec gain = 0	Proportion avec gain > 0	p (test de signe)
103	70	19	14	78.7%	< 1e-4

Interprétation (test de signe) : - test statistiquement significatif ( $p = < 1e-4$ ). Cela indique que les améliorations (gain > 0) sont plus fréquentes que les diminutions (gain < 0), indépendamment de l'amplitude des variations.



#### 4.4 Score initial (Q1) selon l'âge



Interprétation (score initial vs âge) : ce boxplot aide à repérer des niveaux de connaissance initiaux différents selon l'âge, ce qui peut influencer le gain possible ("marge de progression").

Globalement : score initiaux similaires entre les tranches d'âge.

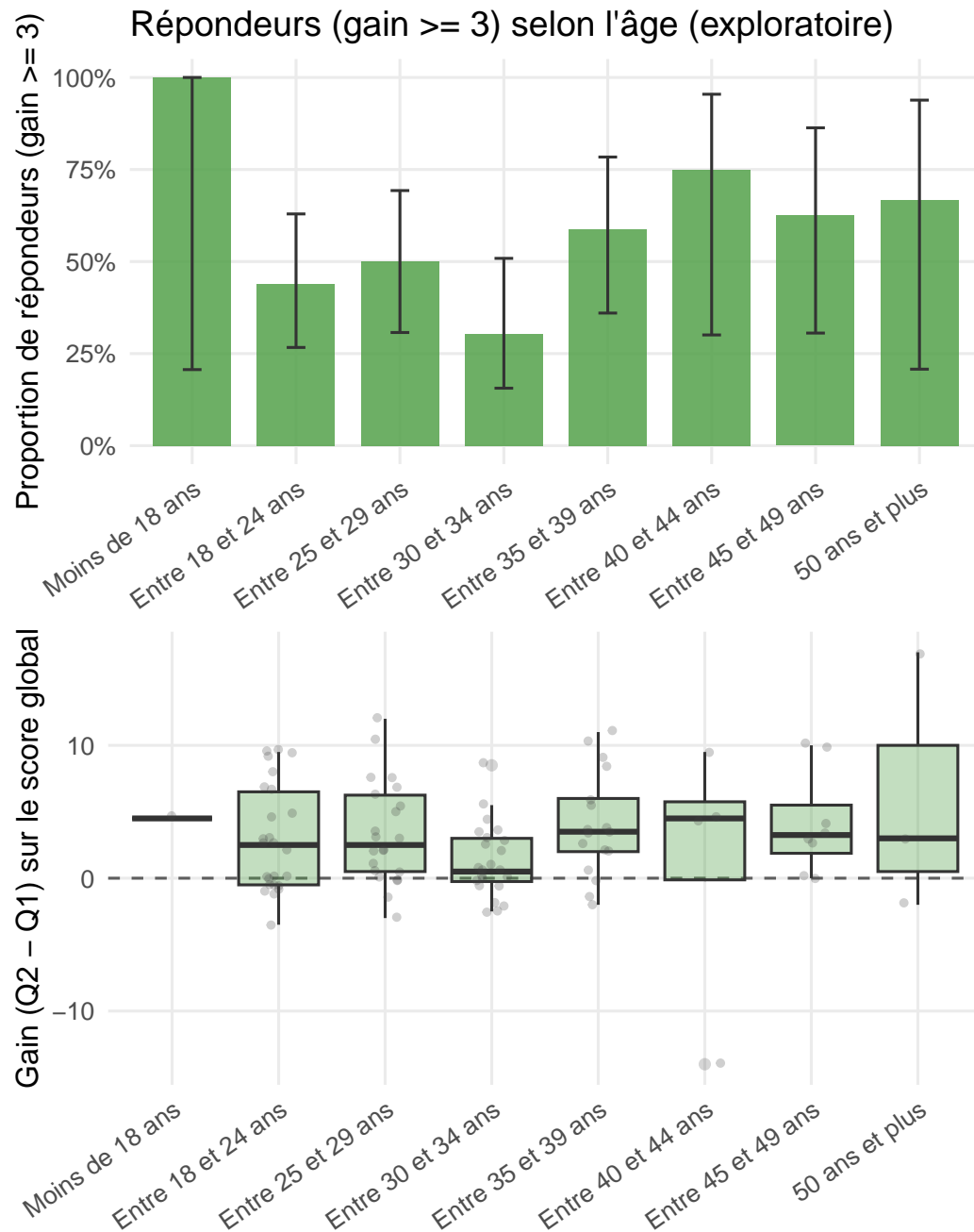
On peut faire un test statistique pour ça (pas obligatoire), qui répond à la question "est-ce que le score initial varie selon l'âge ?"

Interprétation (Kruskal–Wallis, score initial vs âge) : test non statistiquement significatif ( $p = 0.5373$ ). Un résultat significatif suggère des distributions de score initial différentes selon la tranche d'âge.

#### 4.5 Exploration : gain global selon l'âge

Question posée : est-ce que le gain moyen ( $Q2 - Q1$ ) varie selon l'âge des participantes ?

#### 4.5.1 Répondeurs (gain $\geq 3$ ) selon l'âge (exploratoire)



Interprétation (gain vs âge) : Globalement, les gains moyens semblent similaires entre les tranches d'âge.

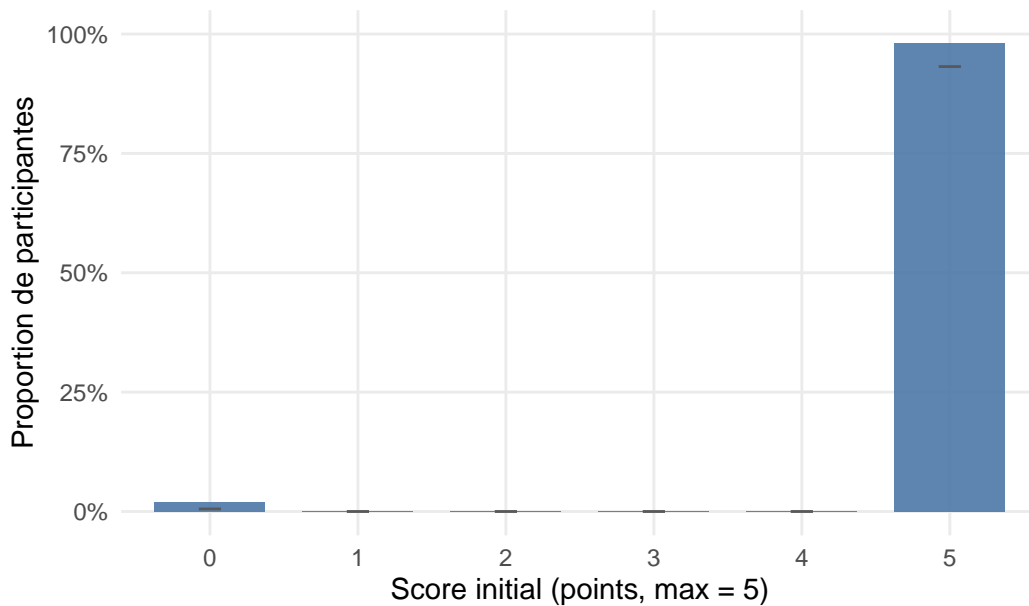
Interprétation (Kruskal-Wallis, gain vs âge) : test non statistiquement significatif ( $p = 0.3518$ ).

Donc pas de différence significative du gain moyen selon l'âge.

#### 4.6 Scores initiaux par question (Q1)

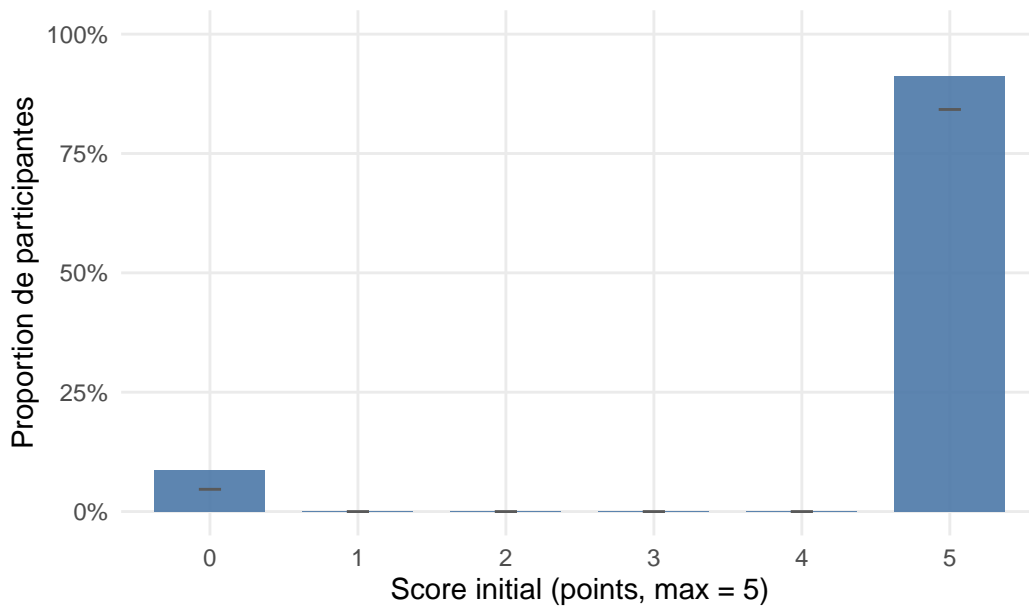
Boire de l'alcool pendant la grossesse nuit au bébé ?

Boire de l'alcool pendant la grossesse nuit au bébé ?

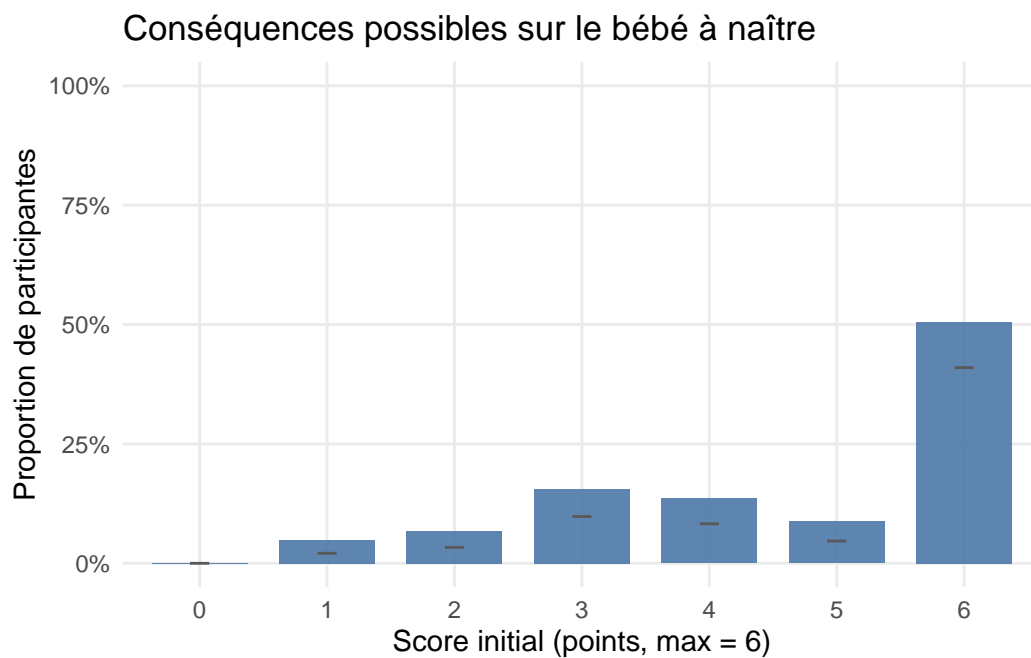


L'alcool bu par la mère arrive jusqu'au bébé ?

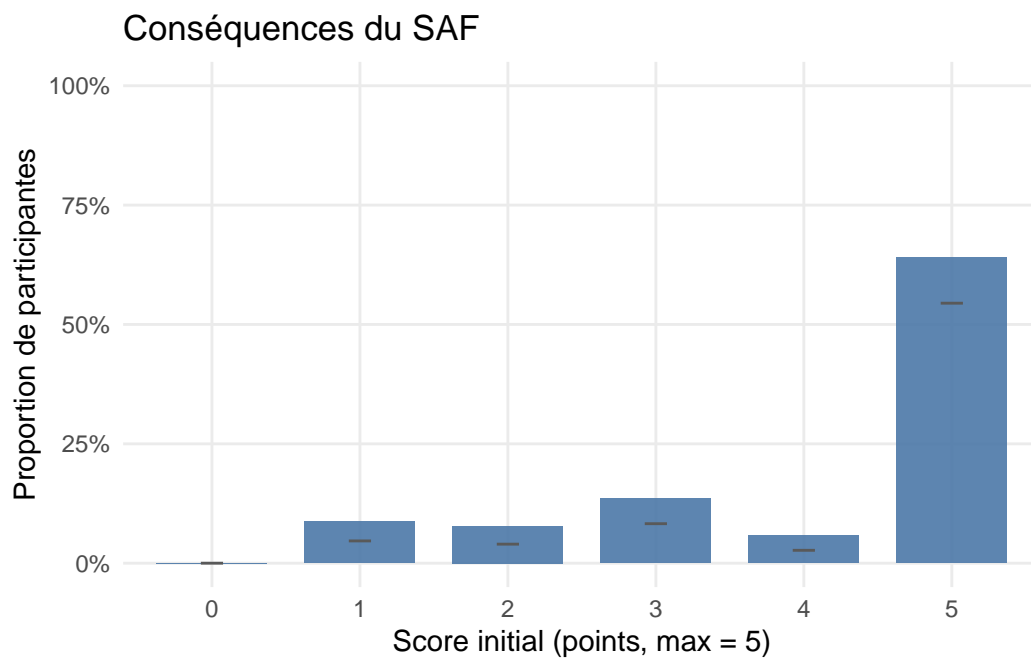
L'alcool bu par la mère arrive jusqu'au bébé ?



### Conséquences possibles sur le bébé à naître

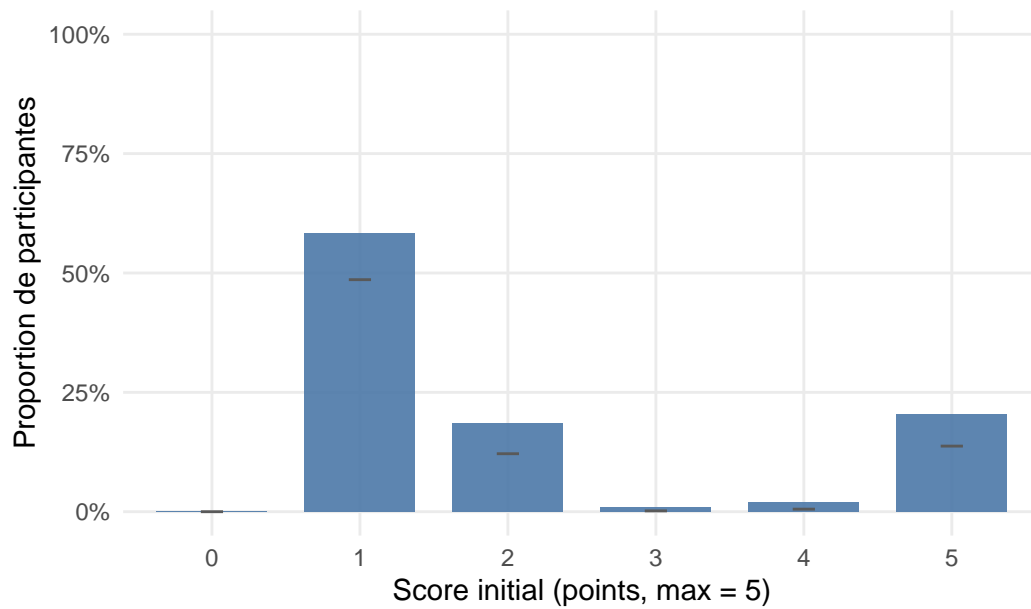


### Conséquences du SAF



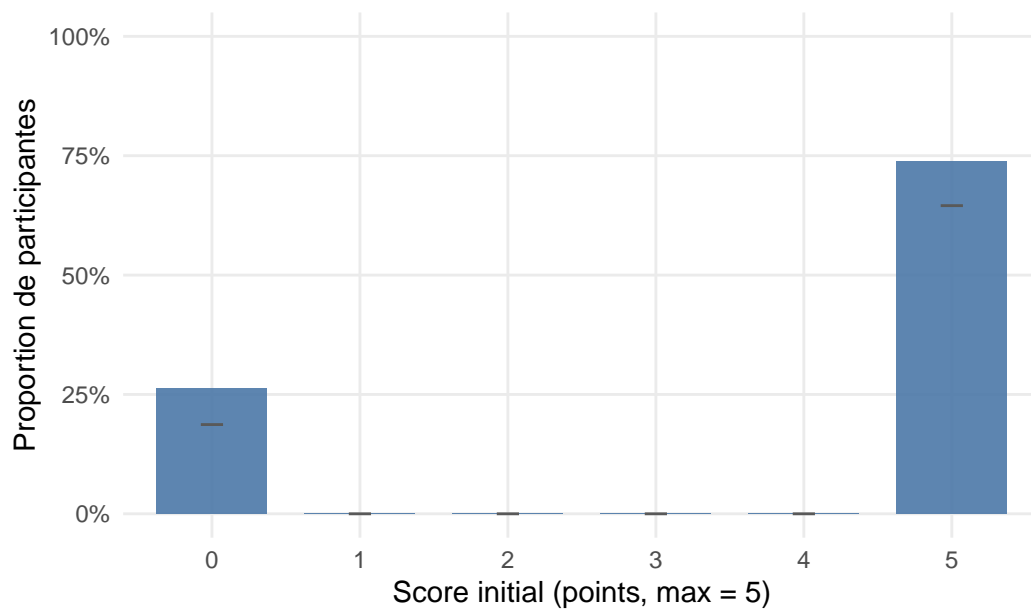
### Âges possibles de détection des troubles

### Âges possibles de détection des troubles

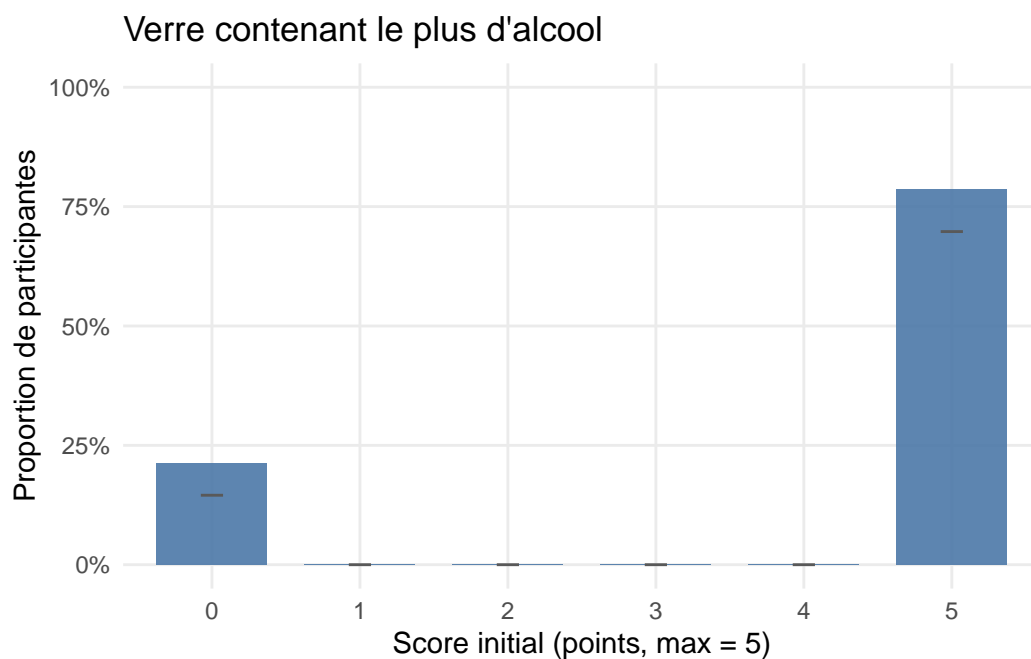


### Quantité d'alcool risquée pendant la grossesse

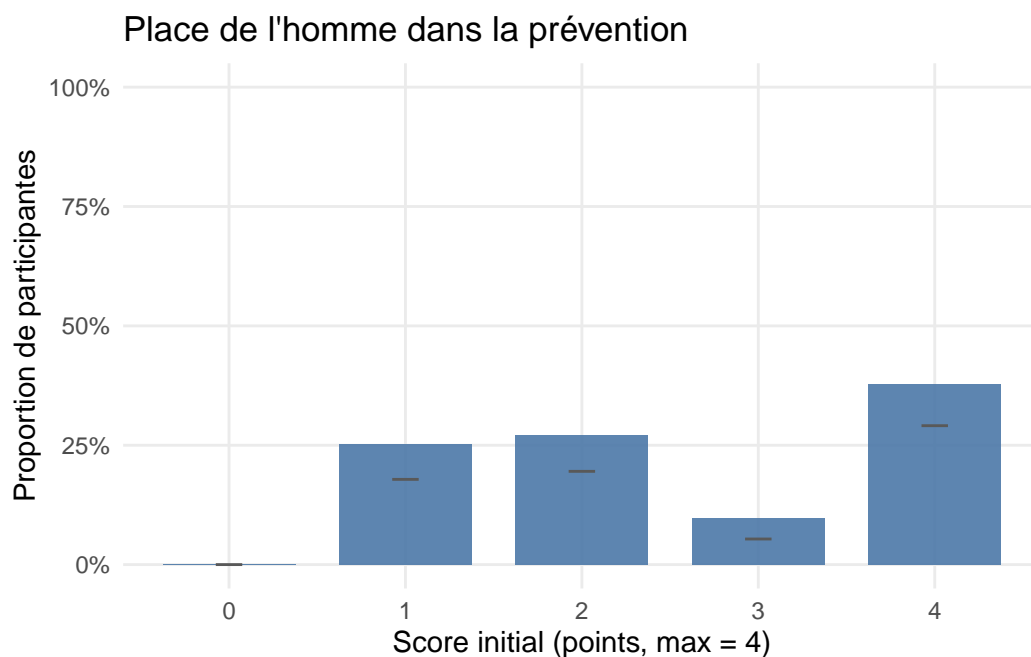
#### Quantité d'alcool risquée pendant la grossesse



### Verre contenant le plus d'alcool



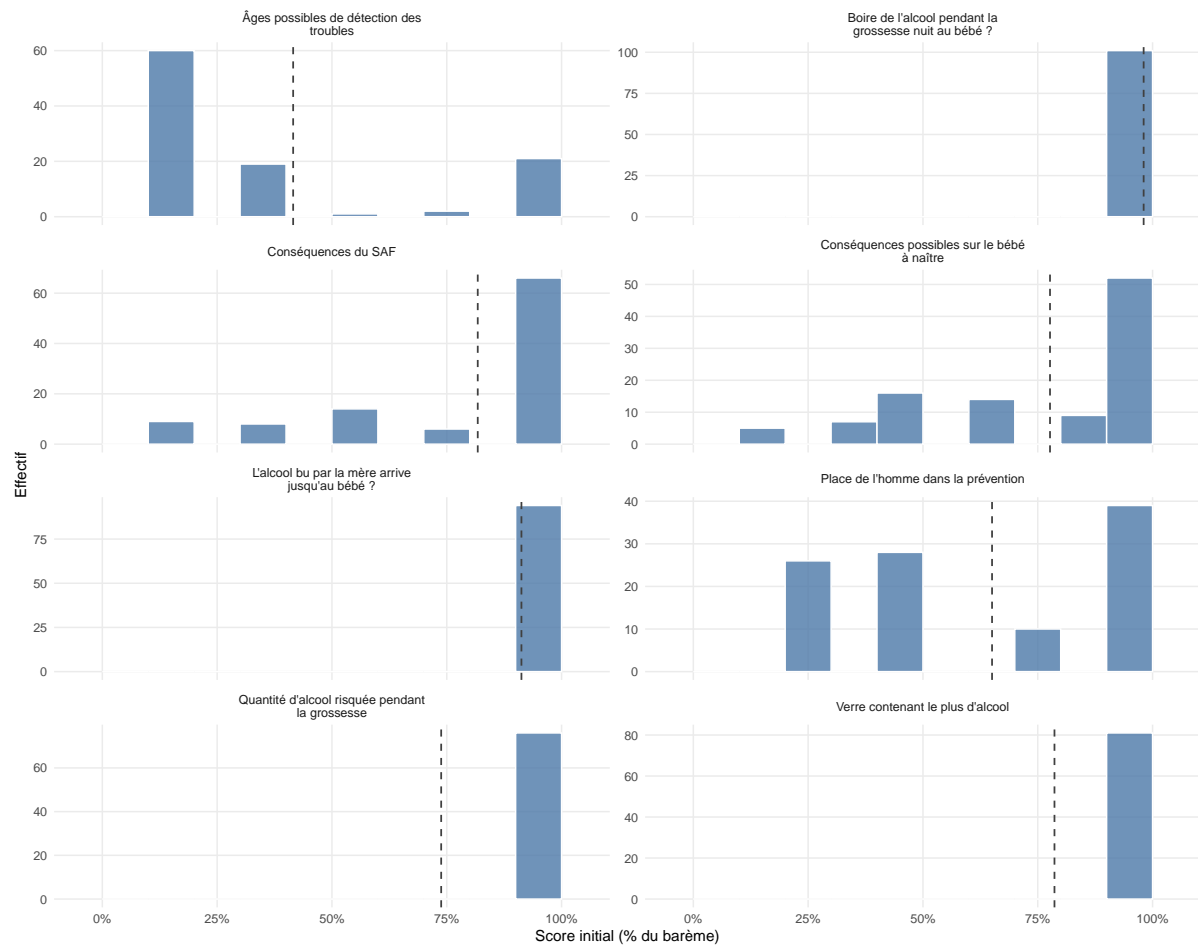
### Place de l'homme dans la prévention



Lorsque les réponses initiales sont bonnes (quasi 100% au score maximal), il existe un probable “effet plafond” de la question, limitant la marge de progression possible.

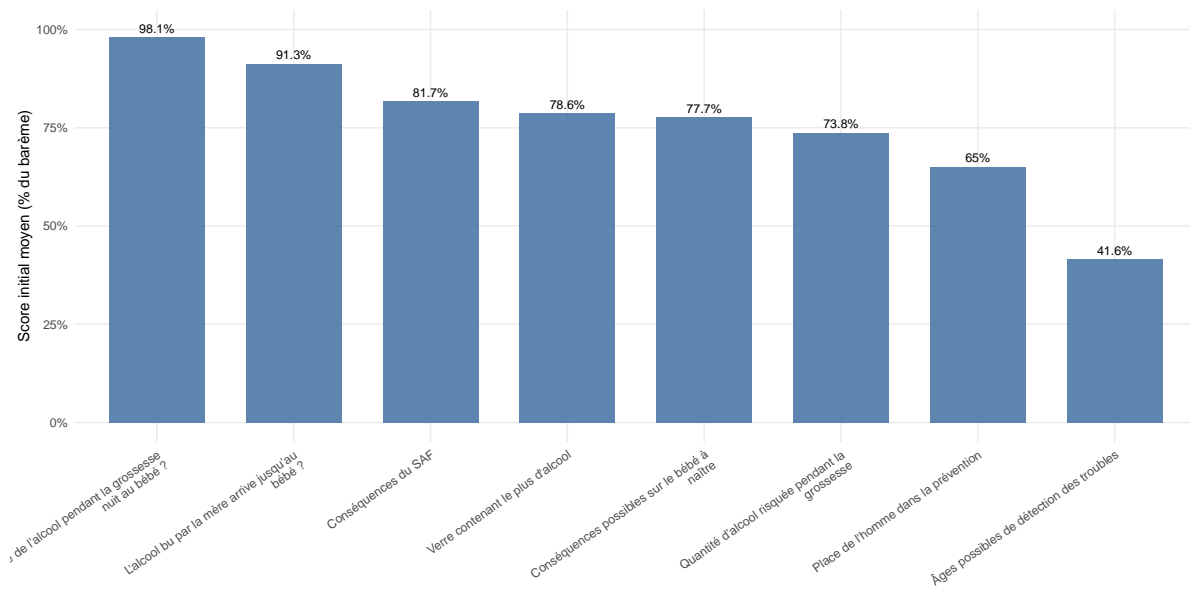
Vue d'ensemble des 8 questions sur un seul graphique :

Distribution des scores initiaux (Q1) par question  
 Ligne pointillée = moyenne. Les questions proches de 100% montrent un effet plafond.





Tous les graphiques sur une seule figure avec un barème normalisé (0–100%) :



## 5 Analyse par question

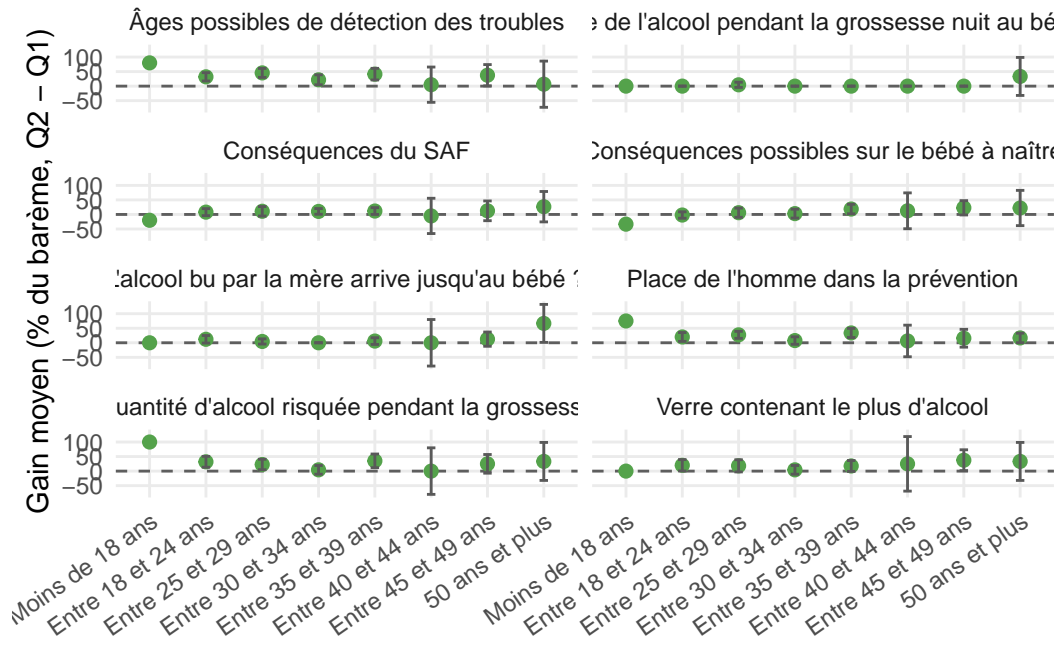
Cette section explore l'évolution **par question** (scores en points et en % du barème) à partir des scores par question déjà calculés plus haut.

Important : ces analyses “par question” sont **exploratoires**. Il n’y a pas de correction pour tests multiples (demande explicite), donc on interprète surtout (i) la **taille** des gains (en points et en % du barème) et (ii) la cohérence avec l’effet plafond, plutôt qu’une liste stricte de p-values  $< 0,05$ .

## 5.1 Résultats par question

Question	Max	n	Moy. Q1	Moy. Q2	Gain moy.	% max Q1	% max Q2
Boire de l'alcool pendant la grossesse nuit au bébé ?	5	103	4.90	5.00	0.10	98.1%	100.0%
Conséquences du SAF	5	103	4.09	4.58	0.50	64.1%	75.7%
Conséquences possibles sur le bébé à naître	6	103	4.66	5.09	0.43	50.5%	62.1%
L'alcool bu par la mère arrive jusqu'au bébé ?	5	103	4.56	4.95	0.39	91.3%	99.0%
Place de l'homme dans la prévention	4	103	2.60	3.43	0.83	37.9%	64.1%
Quantité d'alcool risquée pendant la grossesse	5	103	3.69	4.85	1.17	73.8%	97.1%
Verre contenant le plus d'alcool	5	103	3.93	4.81	0.87	78.6%	96.1%
Âges possibles de détection des troubles	5	103	2.08	3.74	1.66	20.4%	57.3%

Interprétation (résumé par question) : la comparaison des moyennes (en points) et des proportions au score maximal ("full score") permet d'identifier des questions déjà très bien réussies avant (effet plafond) versus celles ayant une marge de progression importante.

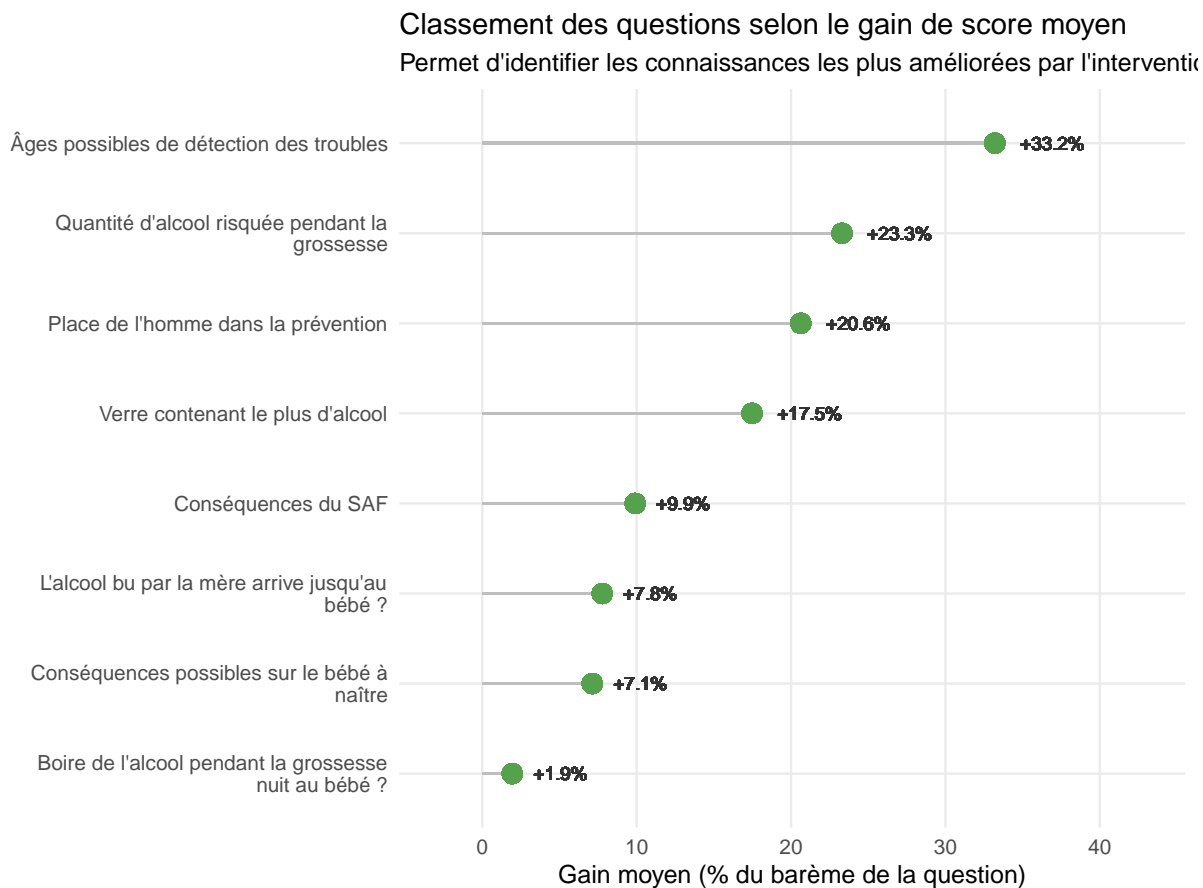


Interprétation (âge × questions) : 2 question(s) montrent une hétérogénéité des gains selon l'âge ( $p < 0.05$ ). Résultat exploratoire, sans correction pour multiplicité.

## 5.2 Synthèse des gains par question

Pour avoir une vue d'ensemble, le graphique suivant représente le gain moyen pour chaque question, exprimé en **pourcentage du barème maximum de la question**.

Les questions sont triées de celle qui a le plus progressé à celle qui a le moins progressé.



Interprétation : Ce graphique synthétise l'impact de l'intervention sur chaque connaissance spécifique.

Les questions sur Âges possibles de détection des troubles et Quantité d'alcool risquée pendant la grossesse ont montré la plus forte progression, tandis que d'autres, probablement déjà bien maîtrisées au départ, ont peu évolué.

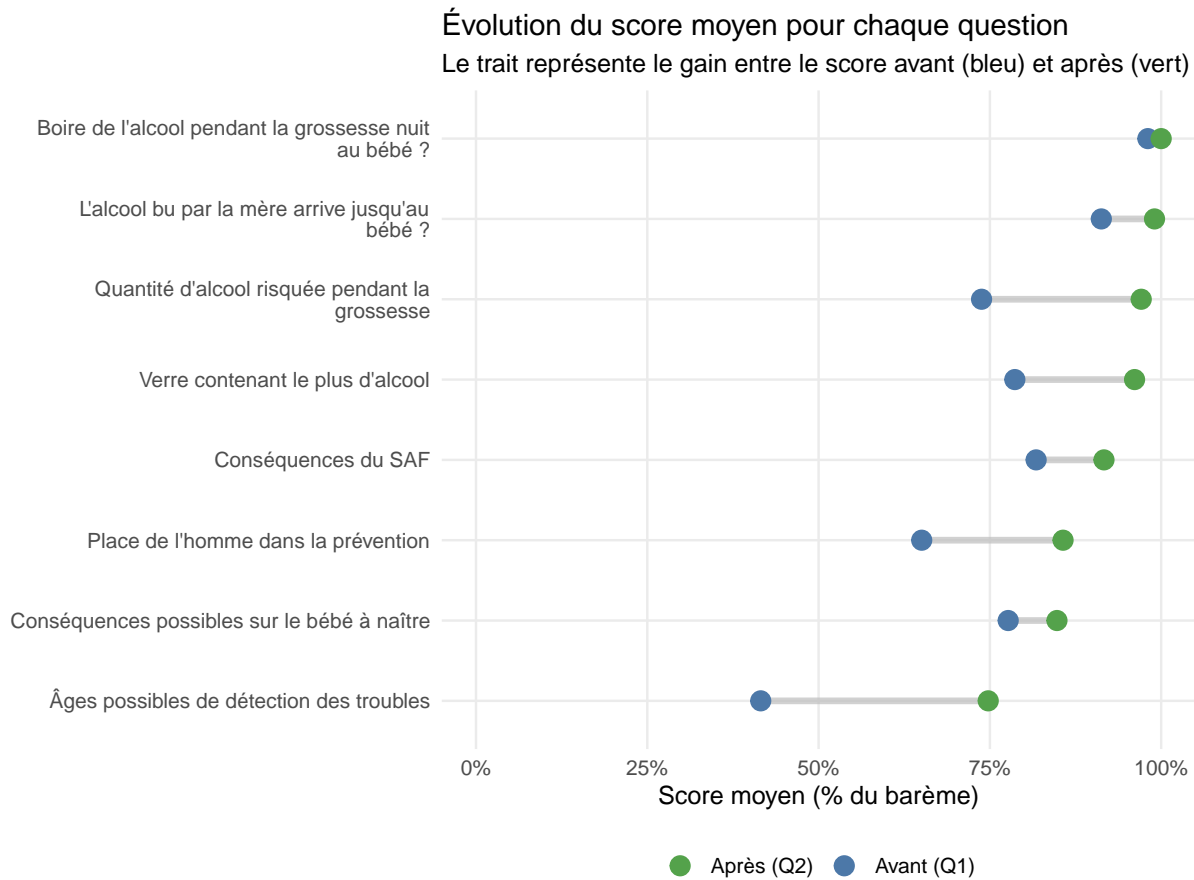
### 5.3 Tests par question (appariés)

Pour chaque item, on teste si le score augmente après intervention (unilatéral : après > avant). Pour les questions à score binaire, cela correspond à une amélioration du taux de “bonne réponse”.

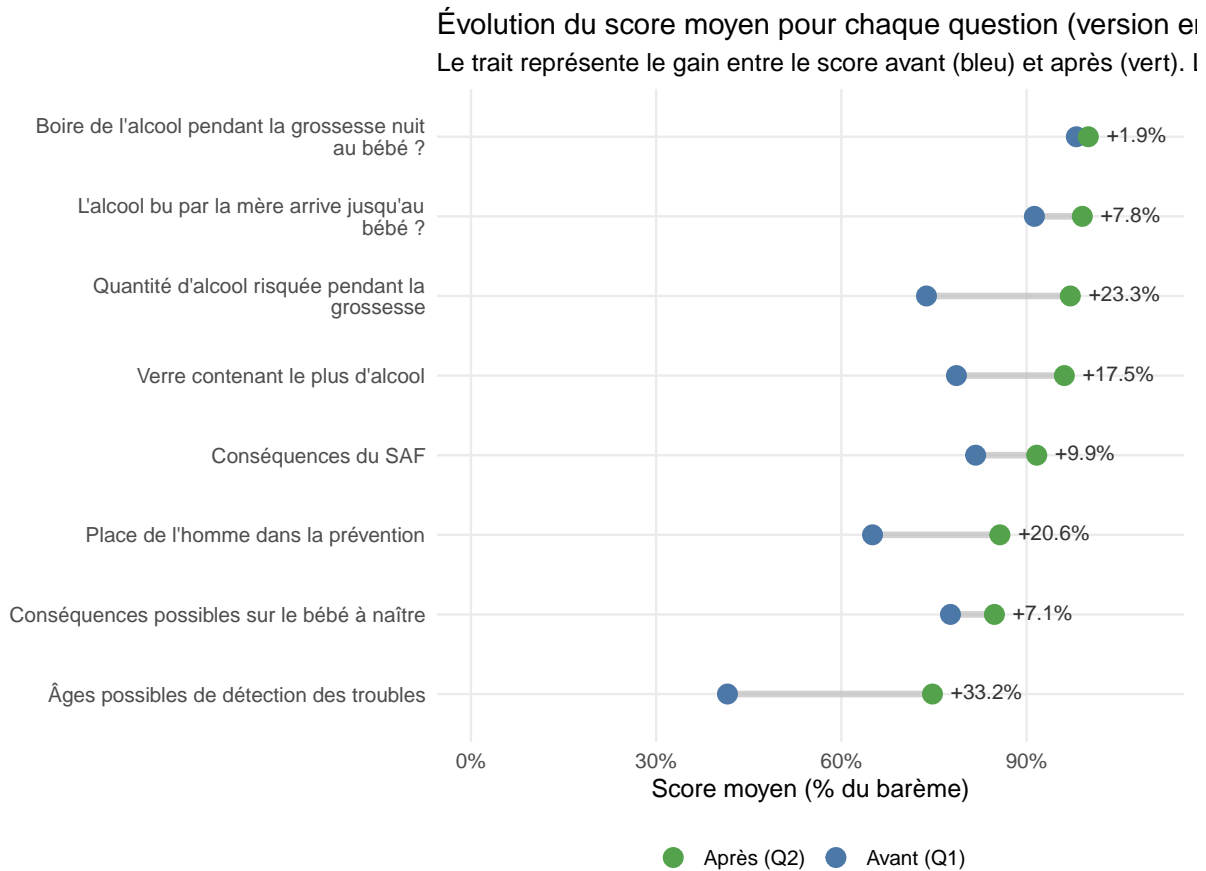
7 question(s) sont significativement améliorées (t-test apparié), résultats exploratoires sans correction pour multiplicité.

Les questions significativement améliorées sont : Conséquences du SAF, Conséquences possibles sur le bébé à naître, L'alcool bu par la mère arrive jusqu'au bébé ?, Place de l'homme dans la prévention, Quantité d'alcool risquée pendant la grossesse, Verre contenant le plus d'alcool, Âges possibles de détection des troubles.

## 5.4 Graphique : score moyen par question (avant vs après)



### 5.4.1 Version enrichie (avec valeur du gain)

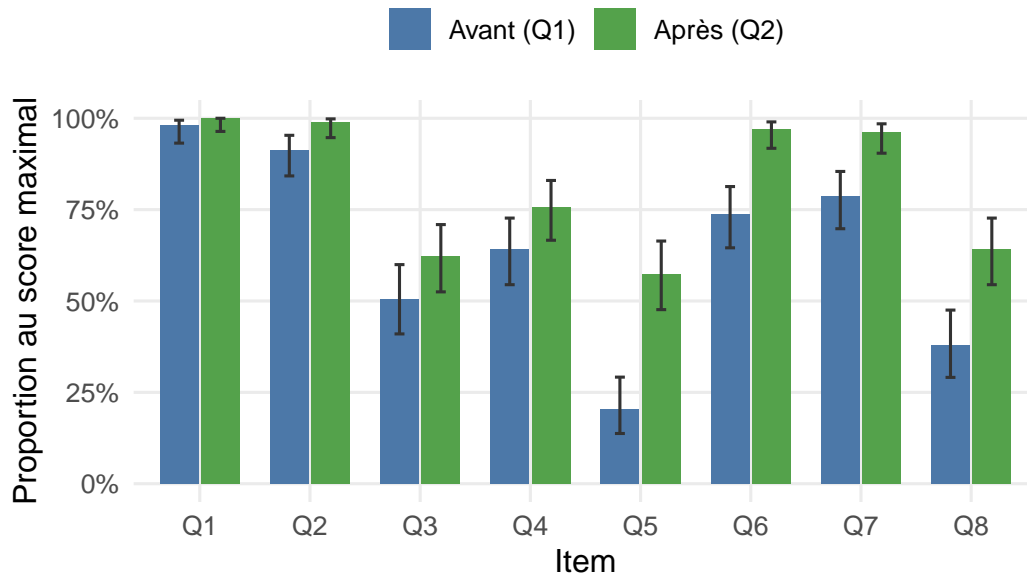


Interprétation (moyennes par question) : les barres représentent la moyenne exprimée en % **du barème** (avec IC95%). Les plus fortes progressions moyennes concernent notamment : Âges possibles de détection des troubles; Quantité d'alcool risquée pendant la grossesse; Place de l'homme dans la prévention.

## 5.5 Autre visualisation de l'amélioration

On visualise bien l'effet plafond sur les questions à score initial élevé.

Effet plafond par item : % au score maximal (avec IC95%)



Interprétation : si l'item est déjà très souvent "au maximum" au Q1, la hausse possible au Q2 est mécaniquement limitée (effet plafond).

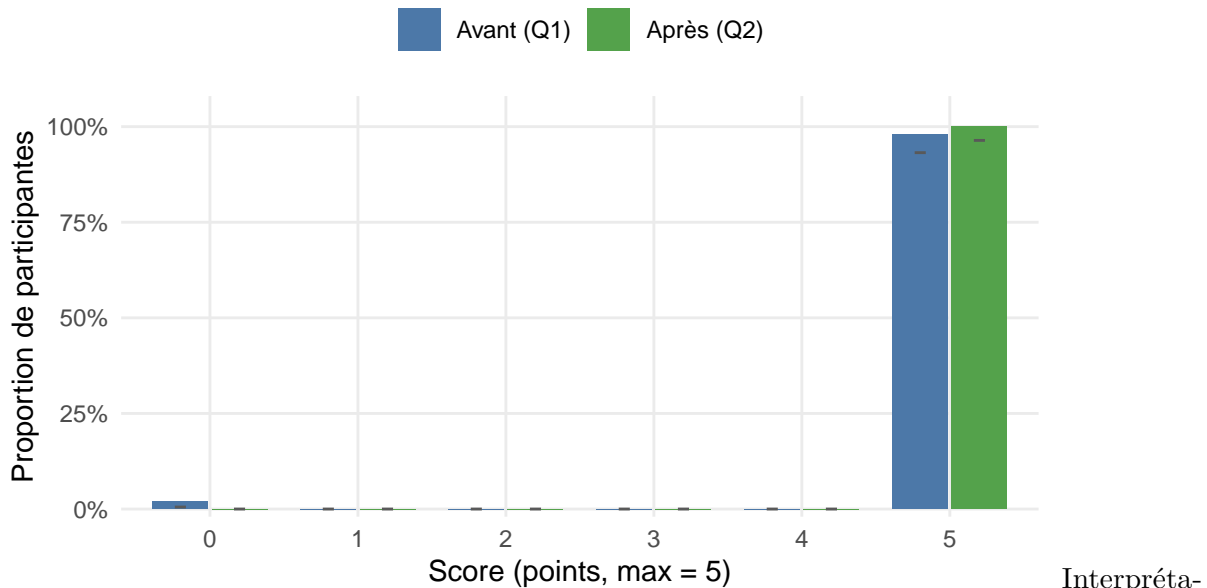


## 5.6 Scores et gains par question (1 graphique par question)

Abscisse = **score en points** (0 à barème max) ; ordonnée = **% de participantes** (avec IC95%).

Boire de l'alcool pendant la grossesse nuit au bébé ?

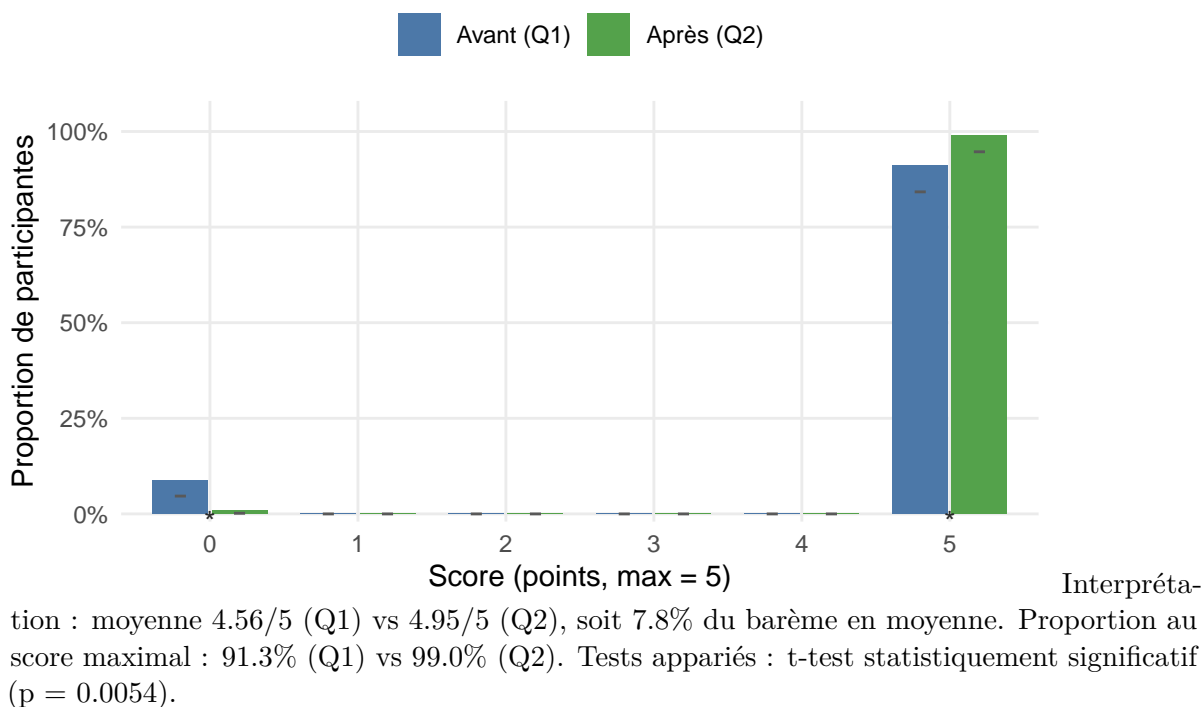
Boire de l'alcool pendant la grossesse nuit au bébé ?



Interprétation : moyenne 4.90/5 (Q1) vs 5.00/5 (Q2), soit 1.9% du barème en moyenne. Proportion au score maximal : 98.1% (Q1) vs 100.0% (Q2). Tests appariés : t-test non statistiquement significatif ( $p = 0.0792$ ).

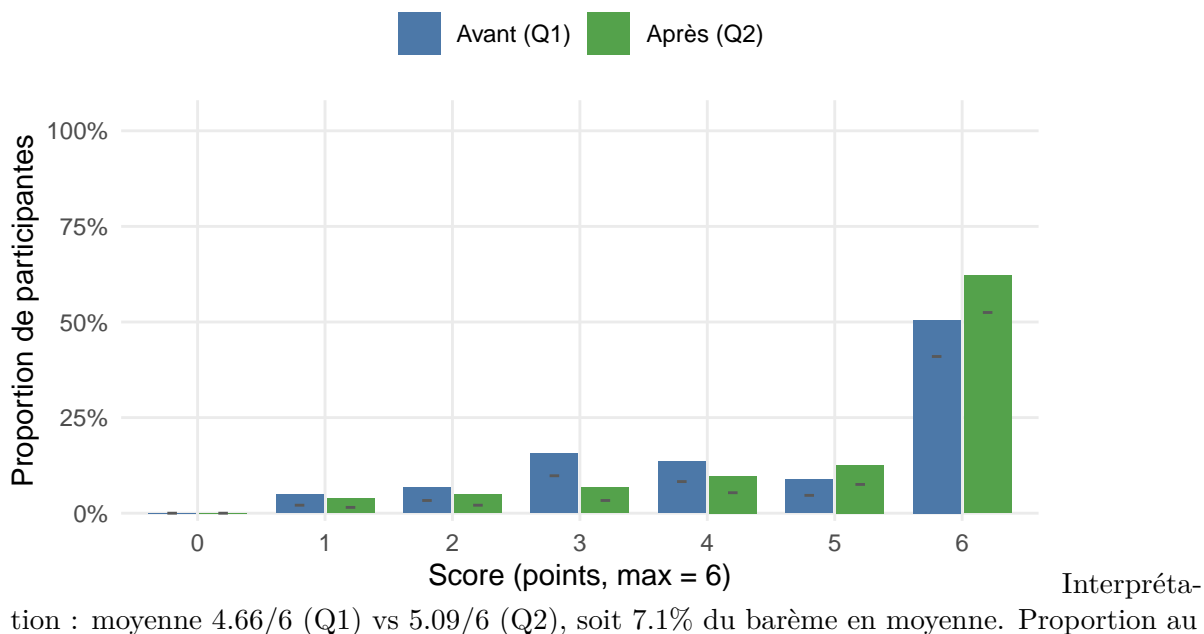
L'alcool bu par la mère arrive jusqu'au bébé ?

## L'alcool bu par la mère arrive jusqu'au bébé ?



## Conséquences possibles sur le bébé à naître

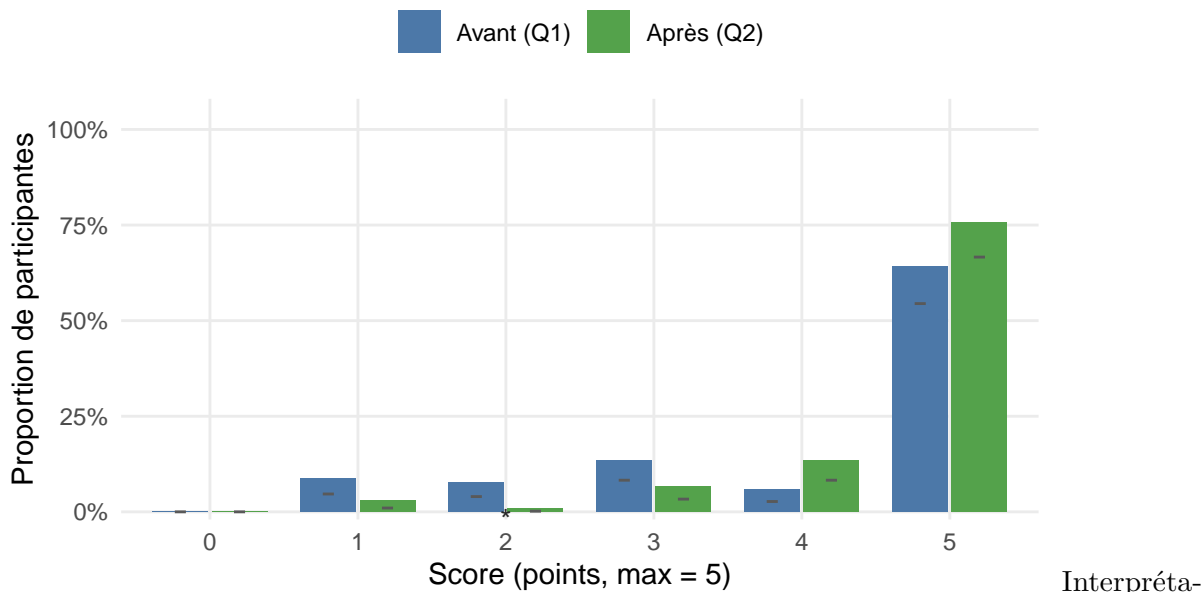
### Conséquences possibles sur le bébé à naître



score maximal : 50.5% (Q1) vs 62.1% (Q2). Tests appariés : t-test statistiquement significatif ( $p = 0.0248$ ).

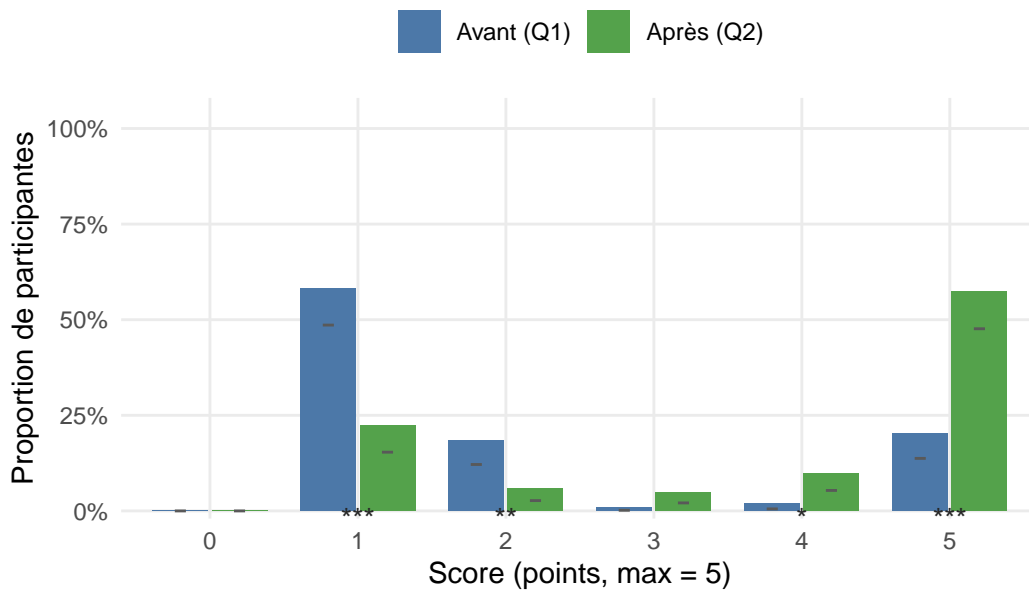
### Conséquences du SAF

#### Conséquences du SAF



### Âges possibles de détection des troubles

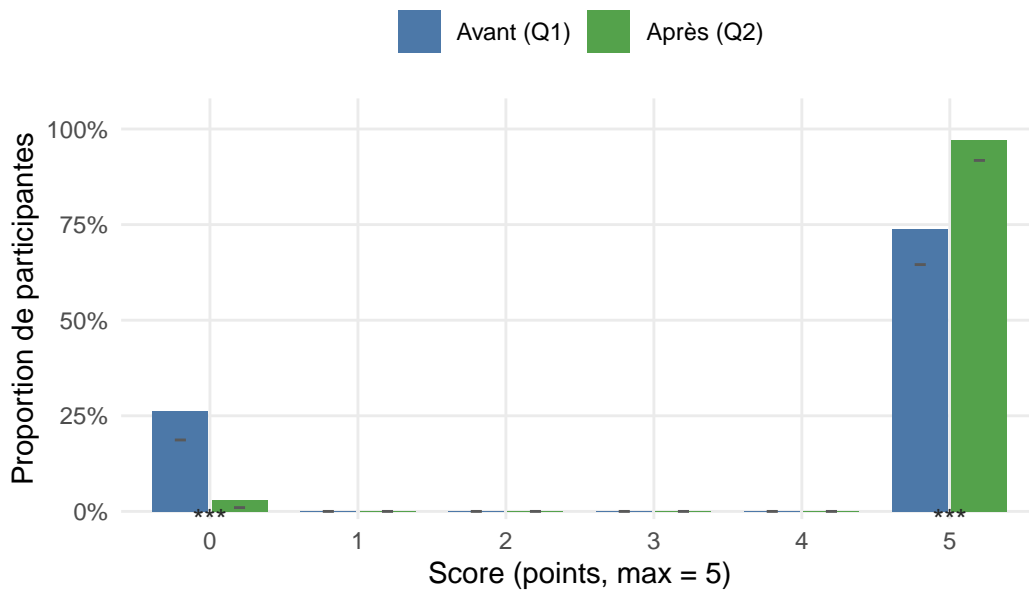
## Âges possibles de détection des troubles



Interprétation : moyenne 2.08/5 (Q1) vs 3.74/5 (Q2), soit 33.2% du barème en moyenne. Proportion au score maximal : 20.4% (Q1) vs 57.3% (Q2). Tests appariés : t-test statistiquement significatif ( $p = < 1e-4$ ).

## Quantité d'alcool risquée pendant la grossesse

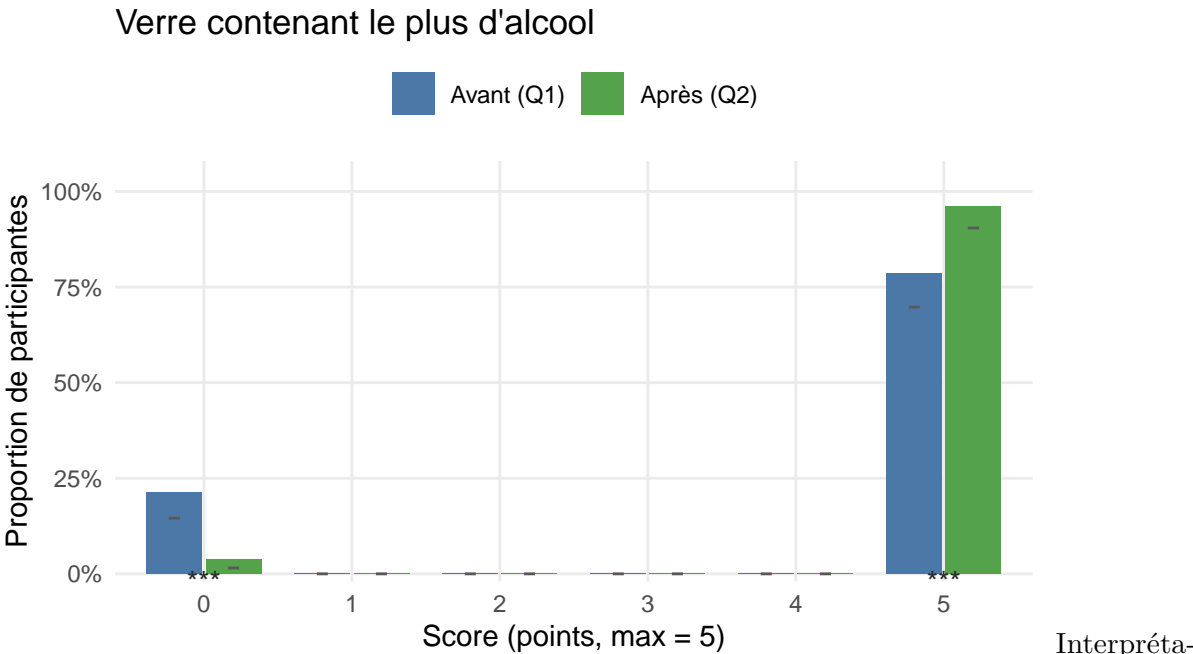
### Quantité d'alcool risquée pendant la grossesse



Interprétation : moyenne 3.69/5 (Q1) vs 4.85/5 (Q2), soit 23.3% du barème en moyenne. Proportion au

score maximal : 73.8% (Q1) vs 97.1% (Q2). Tests appariés : t-test statistiquement significatif ( $p = < 1e-4$ ).

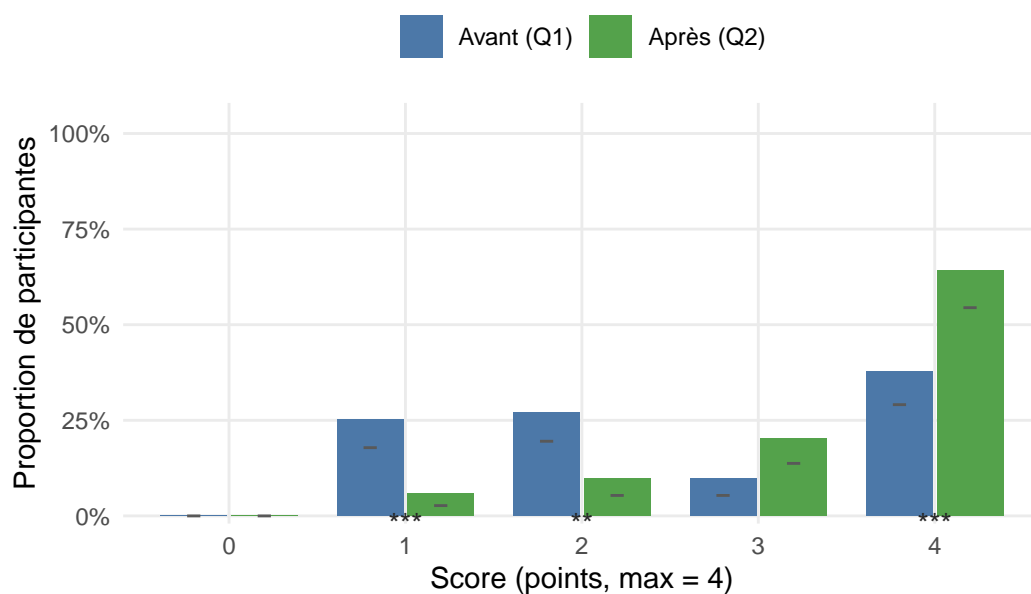
**Verre contenant le plus d'alcool**



Interprétation : moyenne 3.93/5 (Q1) vs 4.81/5 (Q2), soit 17.5% du barème en moyenne. Proportion au score maximal : 78.6% (Q1) vs 96.1% (Q2). Tests appariés : t-test statistiquement significatif ( $p = 0.0001$ ).

**Place de l'homme dans la prévention**

## Place de l'homme dans la prévention

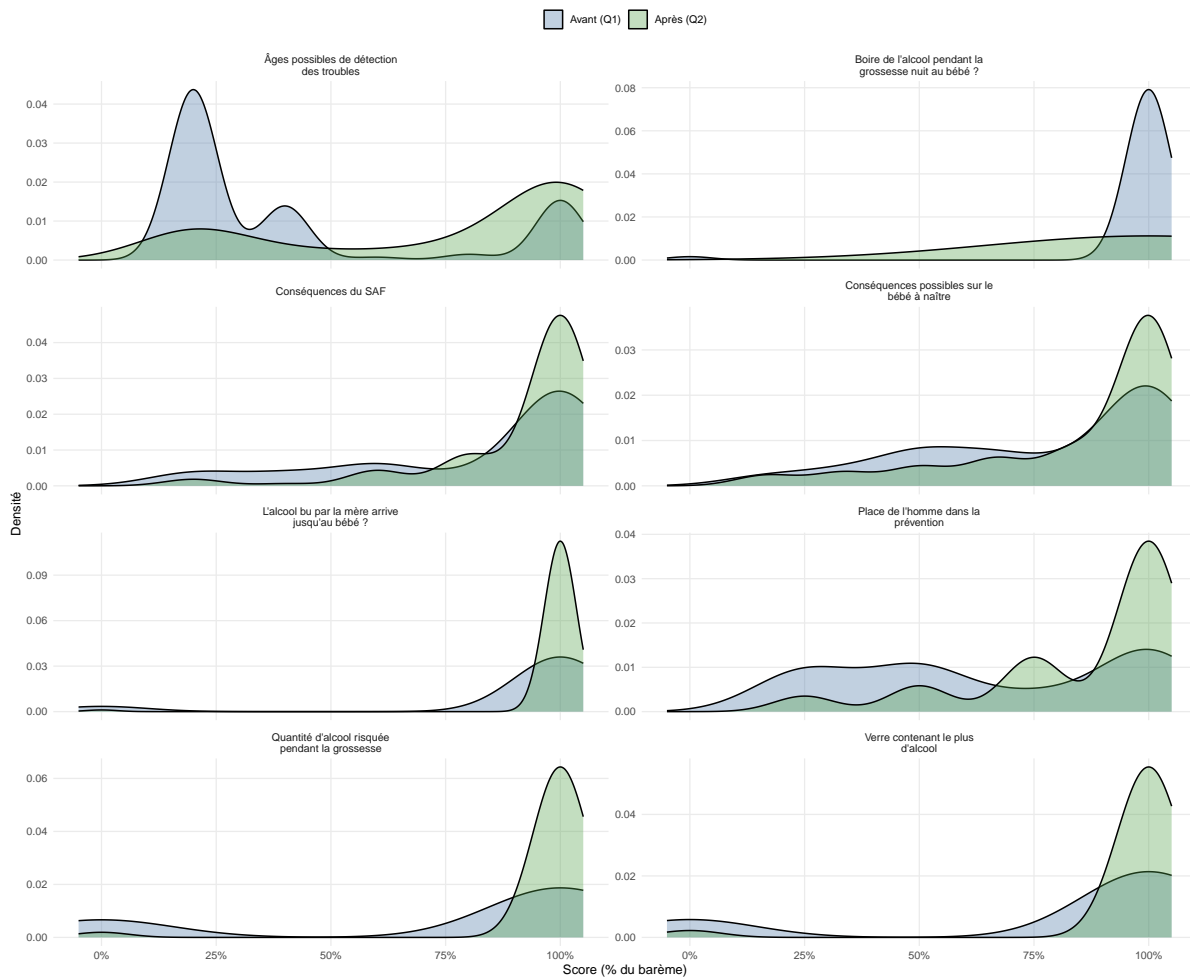


Interprétation : moyenne 2.60/4 (Q1) vs 3.43/4 (Q2), soit 20.6% du barème en moyenne. Proportion au score maximal : 37.9% (Q1) vs 64.1% (Q2). Tests appariés : t-test statistiquement significatif ( $p = < 1e-4$ ).

## 5.7 Vue d'ensemble : distribution Q1 vs Q2

Globalement même principe que les graphiques précédents, mais cette fois on superpose les distributions de scores avant et après intervention pour chaque question.

Évolution des distributions de scores par question (Q1 vs Q2)  
Chaque facette montre le décalage entre avant et après intervention

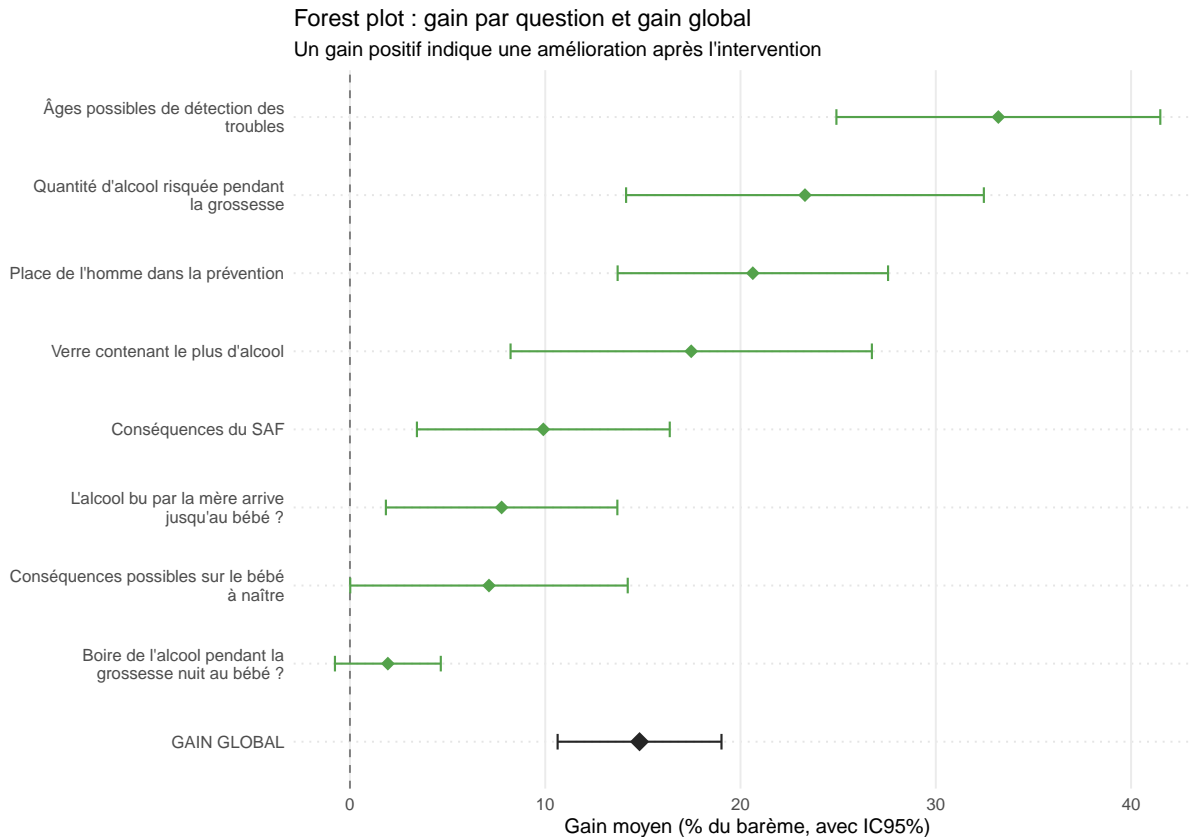


Interprétation : cette vue d'ensemble permet de comparer visuellement les 8 questions sur un seul graphique. Un décalage de la densité vers la droite (vert vs bleu) indique une progression après l'intervention.

## 5.8 Forest plot récapitulatif (gain global + par question)

Gain moyen global et les gains par question (avec IC95%), permettant une comparaison immédiate.

Globalement même principe qu'un des précédents, mais permet d'afficher la significativité des gains (IC95% ne croisant pas zéro).



Le losange noir en bas représente le gain global, tandis que les points verts montrent les gains par question.

Les questions dont l'IC95% ne croise pas zéro sont celles où l'amélioration est statistiquement significative.

## 6 Biais de sélection potentiel : Q1 appariées vs Q1 non appariées

### ! Important

Pourquoi on compare Q2 à la moyenne Q1 *appariée* (et pas à la moyenne Q1 "totale") :

- Pour mesurer un effet "avant/après", il faut comparer **les mêmes personnes** au temps 1 et au temps 2.

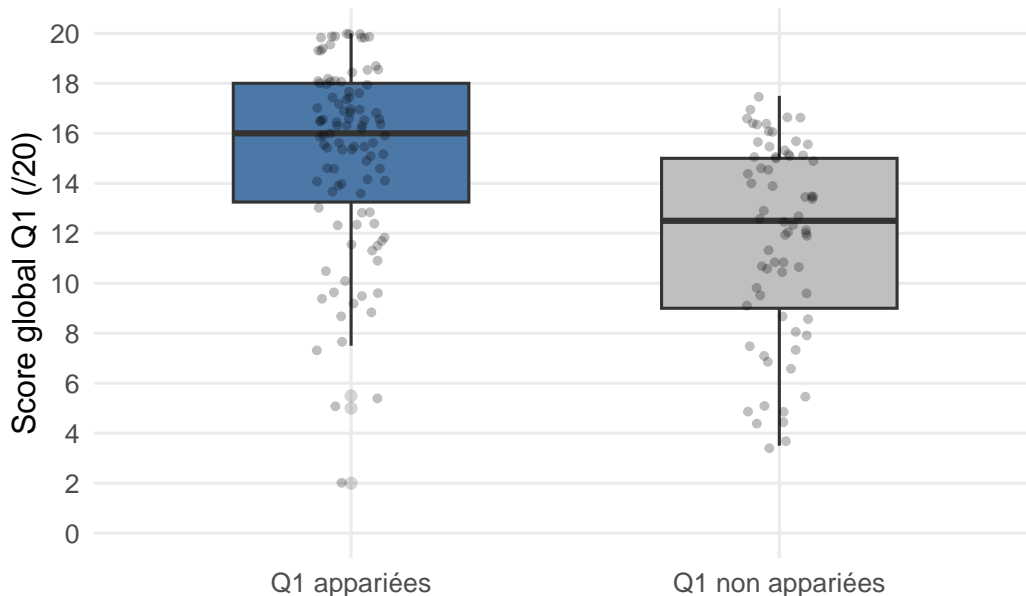


- La moyenne Q2 est calculée uniquement chez les participantes qui ont un Q2 (donc les **appariées**).
- Si on compare la moyenne Q2 (appariées) à la moyenne Q1 “totale” (appariées + non appariées), on compare en réalité **deux groupes différents**. La différence mélange alors l’effet potentiel de l’information avec un **biais de sélection** (par exemple : les participantes qui répondent au Q2 peuvent être, au départ, plus informées ou plus motivées).
- Dans ces données, cela se voit : la moyenne Q1 avant appariement (**13,88/20**) est différente de la moyenne Q1 chez les appariées (**15,26/20**). Utiliser la moyenne Q1 “totale” pour estimer le gain donnerait donc un gain artificiellement plus grand, qui ne correspond pas à une comparaison avant/après sur les mêmes personnes.
- On pourrait aussi être tenté d’utiliser la moyenne Q2 “totale” (toutes les répondantes au Q2, **18,19/20** sur 106 réponses, min = 4,50, max = 20,00), mais cela ne règle pas le problème : le calcul d’un gain nécessite de disposer de Q1 et Q2 **pour la même personne**.

En gros, on compare

- les patientes ayant répondu au Q1 avant appariement
- les patientes ayant répondu au Q1 après appariement (donc ayant aussi répondu au Q2)
- pour vérifier si ces deux groupes ont des scores initiaux (Q1) similaires ou différents.

### Comparaison des scores Q1 : appariées vs non appariées



Le score initial dans le groupe “apparié” est différent.

Utiliser la moyenne Q1 “totale” pour estimer un gain (en la comparant à la moyenne Q2) serait **non valide**, car cela comparerait des populations différentes donc induirait un biais de sélection.

! Pourquoi on compare Q2 à la moyenne Q1 *appariée* (et pas à la moyenne Q1 “totale”)

- Pour mesurer un effet “avant/après”, il faut comparer **les mêmes personnes** au temps 1 et au temps 2.
- La moyenne Q2 est calculée uniquement chez les participantes qui ont un Q2 (donc les **appariées**).
- Si on compare la moyenne Q2 (appariées) à la moyenne Q1 “totale” (appariées + non appariées), on compare en réalité **deux groupes différents**. La différence mélange alors l’effet potentiel de l’information avec un **biais de sélection** (par exemple : les participantes qui répondent au Q2 peuvent être, au départ, plus informées ou plus motivées).
- Dans ces données, cela se voit : la moyenne Q1 avant appariement (**13,88/20**) est différente de la moyenne Q1 chez les appariées (**15,26/20**). Utiliser la moyenne Q1 “totale” pour estimer le gain donnerait donc un gain artificiellement plus grand, qui ne correspond pas à une comparaison avant/après sur les mêmes personnes.
- On pourrait aussi être tenté d’utiliser la moyenne Q2 “totale” (toutes les répondantes au Q2, **18,19/20** sur 106 réponses, min = 4,50, max = 20,00), mais cela ne règle pas le problème : le calcul d’un gain nécessite de disposer de Q1 et Q2 **pour la même personne**.

! Ce que cela signifie concrètement

Les tests et le graphique confirment que les participantes qui ont répondu au Q2 (“appariées”) avaient déjà un **score initial significativement plus élevé**.

- **Score Q1 moyen (appariées) : 15,26**
- **Score Q1 moyen (non appariées) : 11,81**
- **Score Q1 moyen (total) : 13,88**

Il serait donc **méthodologiquement faux** de comparer la moyenne de *tous* les Q2 à la moyenne de *tous* les Q1. La seule méthode correcte est l’analyse **appariée**, qui mesure l’évolution au sein de chaque personne ayant complété les deux questionnaires.

Interprétation possible (à discuter) : si les participantes non appariées avaient eu un score initial très élevé, on pourrait suspecter qu’elles avaient moins ressenti le besoin de répondre au second questionnaire.

**Ici, c’est plutôt l’inverse** (les appariées ont un score initial moyen plus élevé), ce qui suggère plutôt que les participantes les plus informées (ou les plus motivées) ont davantage répondu au Q2.

## 7 Discussion

### 7.1 Résultats principaux

#### Les 3 conclusions clés

1. **L'intervention a fonctionné** : amélioration statistiquement significative des connaissances, avec un gain moyen de **2.97 points** (IC95% : [2.13 ; 3.81], d de Cohen = 0.69, effet moyen). Le message principal est positif.
2. **L'effet est modéré par un effet plafond** : le score de départ était déjà élevé (15,26/20), limitant la marge de progression. Le gain relatif moyen (34.9% de la marge disponible) est plus informatif que le gain brut.
3. **L'effet est homogène** : aucune différence significative selon l'âge (Kruskal-Wallis :  $p = 0.3518$ ). L'intervention bénéficie à toutes les participantes de manière comparable.

Sur 103 participantes appariées, le score moyen passe de **15.26/20** à **18.22/20**, soit un gain moyen de **2.97 point(s)** (t-test apparié :  $p = < 1e-4$ ). En pratique, **50/103** participantes (**48.5%**) ont un gain  $\geq 3$  points. L'analyse par question (exploratoire) montre que les plus fortes progressions concernent : Âges possibles de détection des troubles; Quantité d'alcool risquée pendant la grossesse; Place de l'homme dans la prévention.

### 7.2 Limites méthodologiques

1. **Biais de sélection (attrition)** : les participantes appariées avaient un score initial plus élevé que les non appariées, ce qui suggère un biais de sélection par motivation/engagement. L'analyse avant/après reste valide sur les paires, mais la généralisation doit être prudente. L'estimation d'un gain "avant/après" doit reposer sur les participantes ayant les deux mesures ; la moyenne Q2 "totale" (18,19/20 sur 106 réponses) reste descriptive et ne peut servir à estimer un gain.
2. **Appariement indirect** : l'appariement repose sur l'âge et la ville, sans identifiant unique. Toute erreur d'appariement tend à "diluer" l'effet mesuré.
3. **Absence de groupe contrôle** : sans groupe comparable, on ne peut pas attribuer formellement l'amélioration à la seule intervention. C'est une limite classique des études avant-après.
4. **QCM = connaissances, pas comportements** : mieux répondre au QCM ne garantit pas un changement de comportement, même si c'est un premier indicateur.

### 7.3 Recommandations pour une future étude

1. **Identifiant unique anonyme** pour relier Q1 et Q2 de façon certaine et limiter les pertes au suivi.
2. **Groupe contrôle randomisé** (information différée à un autre groupe) pour isoler l'effet de l'intervention.
3. **Questionnaire plus discriminant** : questions plus difficiles pour limiter l'effet plafond et mieux discriminer les niveaux de connaissance.
4. **Critères de jugement élargis** : mesurer les intentions de changement, l'auto-efficacité, et idéalement des indicateurs de comportement, avec un suivi à distance (3-6 mois).

### 7.4 Version thèse (paragraphe)

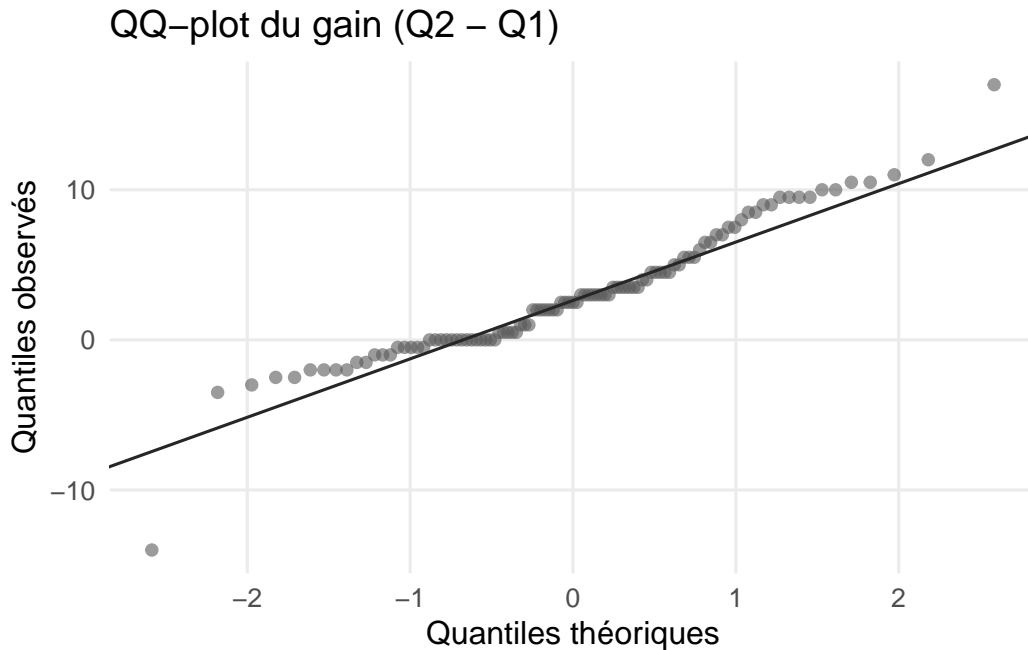
Dans cette étude avant-après ( $n = 103$  paires), le score moyen au QCM augmente de **15.26/20** avant information à **18.22/20** après information, soit un gain moyen de **2.97 point(s)** (d de Cohen = 0.69, effet moyen). Les tests appariés concluent à une amélioration statistiquement significative (t-test unilatéral :  $p = < 1e-4$ ). L'intervalle de confiance bilatéral du gain moyen est [**2.13 ; 3.81**]. Au niveau individuel, **50/103** participantes (**48.5%**) atteignent un gain  $\geq 3$  points. Le gain relatif moyen (rapporté à la marge de progression théorique) est de **34.9%**, ce qui permet de nuancer l'effet plafond observé sur le gain brut. L'analyse exploratoire par régression logistique (ajustée sur l'âge et la langue) confirme que le facteur principal est le score initial (effet plafond), sans effet clair de l'âge ni de la langue. L'analyse par question indique que les plus fortes progressions concernent : Âges possibles de détection des troubles; Quantité d'alcool risquée pendant la grossesse; Place de l'homme dans la prévention.

Plusieurs limites peuvent influencer l'interprétation : l'appariement repose sur des informations indirectes (âge, ville), l'absence de groupe contrôle empêche d'attribuer formellement l'évolution au seul message d'information, et un QCM mesure surtout des connaissances à court terme. Pour améliorer l'évaluation dans une étude future, il serait souhaitable d'utiliser un identifiant unique, d'ajouter un groupe contrôle, de renforcer la sensibilité du questionnaire et d'intégrer des critères plus proches de l'objectif clinique, avec un suivi à distance pour évaluer la persistance de l'effet.

## 8 Annexes techniques (détails statistiques)

Cette section regroupe des compléments “techniques” utiles pour un lecteur souhaitant vérifier certains points méthodologiques, sans alourdir l’interprétation principale.

### 8.1 Diagnostic visuel (gain global)



Interprétation : ce QQ-plot donne une idée de l’éloignement à la normalité. L’analyse principale ne repose pas uniquement sur la normalité car le bootstrap du gain moyen est également rapporté.

Globalement, la distribution semble normale. On aurait pu ne même pas faire de *bootstrap* ! C’est peut être un excès de prudence, mais cela ne nuit pas à l’analyse.

A la question “avez vous fait un test de normalité (type Shapiro-Wilk) ?” : non : ce n’est pas utile.

- Les tests de normalité reposent sur les hypothèses :
- $H_0$  : on ne peut pas prouver que la distribution n’est pas normale
- $H_1$  : la distribution n’est pas normale
- Or, avec un échantillon de taille modérée (ici  $n = 103$ ), on n’a pas beaucoup de puissance pour détecter une déviation modérée à la normalité.

- Donc si la p-value est  $> 0.05$ , on ne peut pas conclure que la distribution est normale (simplement qu'on n'a pas pu prouver le contraire).
- Principe simple : l'absence de preuve n'est pas la preuve de l'absence !!

## 8.2 Bootstrap

Ici, on a réalisé un bootstrap non paramétrique pour estimer l'incertitude (IC95%) sur la moyenne des scores avant, après, et du gain.

Globalement, ce n'était pas strictement nécessaire (la distribution semble normale), mais c'est une précaution supplémentaire.

Le bootstrap est une méthode de rééchantillonnage permettant d'estimer la distribution d'un estimateur (ici, la moyenne) sans s'appuyer sur des hypothèses paramétriques fortes (comme la normalité).

Principe :

- on dispose d'un échantillon de taille  $n$  (ici, les 103 participantes appariées) ;
- on tire  $B$  échantillons bootstrap de taille  $n$  en repiquant avec remise parmi les  $n$  observations initiales ;
- pour chaque échantillon bootstrap, on calcule l'estimateur d'intérêt (ici, la moyenne des scores avant, après, et du gain) ;
- la distribution des  $B$  valeurs obtenues permet d'estimer l'incertitude de l'estimateur (par exemple, un intervalle de confiance).
- on peut également visualiser la distribution des valeurs bootstrap pour mieux comprendre l'incertitude.
- et ainsi utiliser des tests paramétriques (type t-test) basés sur cette distribution bootstrap.

Interprétation (bootstrap, 1000 répliques) : l'IC95% bootstrap du gain moyen est [**2.16, 3.74**]. Cet intervalle est compatible avec un gain moyen  $> 0$  et montre que le gain moyen est plutôt autour de 2-3 points.

### 8.2.1 Densités (avant bootstrap) et densités des moyennes (bootstrap)

Le premier graphique représente la distribution des **scores individuels observés**.

Le second représente la distribution bootstrap des **moyennes** (après 1000 repiquages). La ligne verticale pointillée correspond à la moyenne observée.

On observe bien que la distribution devient "normale".

