VALIDITÉ DES DONNÉES : APPORT DES SOURCES MULTIPLES POUR L'IDENTIFICATION ET L'ÉTUDE DES CANCERS INCIDENTS

Sofiane KAB

26/05/2026

9e Journée Scientifique Constances & Gazel

01.

Cancer - enjeu majeur de santé publique

OMS — GLOBOCAN 2020

10 M
Décès liés au cancer en 2020 (OMS)

1 / 6
Décès dans le monde dûs au cancer

42 %
Des nouveaux cas mondiaux
→ 4 cancers ciblés

Poumon
2,2 M cas | 1,8 M décès
Cancer le plus meurtrier

Colorectal
1,9 M cas | 935 000 décès
2e cause mortalité mondiale

Sein
2,3 M cas | 685 000 décès
1re cause mortalité féminine

Prostate
1,4 M cas | 375 000 décès
1er cancer masculin

La qualité des données conditionne la validité des résultats

Question : comment une même association change selon la source utilisée ?

- **Pathologies d'intérêt :** cancers du poumon, colorectal, sein et prostate
- **Expositions :** tabac, obésité, alcool

02.

Identification et validation des cancers dans Constances

Constances : Identification via questionnaires

Questionnaire médecin d'inclusion

10. *Cancer :* ☐ 1 Oui ☐ 2 Non

Si oui	Age au diagnostic
☐ 1 Sein	ans
☐ 1 Utérus Col	ans
☐ 1 Utérus Corps	ans
☐ 1 Ovaire	ans
☐ 1 Thyroïde	ans
☐ 1 Broncho-pulmonaire	ans
☐ 1 Prostate	ans
☐ 1 Colon/rectum	ans
☐ 1 Autre(s) cancer(s), précisez	

Questionnaire auto-administré aux suivis annuels

Cancer ..... ☐ 43

↘ Précisez la localisation :

☐ 1 Côlon-rectum	☐ 1 Poumon	☐ 1 Prostate	☐ 1 Sein
☐ 1 Col de l'utérus	☐ 1 Utérus (endomètre)		☐ 1 Os
☐ 1 Mélanome	☐ 1 Vessie	☐ 1 Testicule	☐ 1 Rein
☐ 1 Thyroïde	☐ 1 Estomac	☐ 1 Pancréas	☐ 1 Foie
☐ 1 Cerveau	☐ 1 Leucémie	☐ 1 Lymphome	☐ 1 Autre

Constances : Identification via SNDS

Cartographie CNAM

Cancers	Identification
Cancer du sein actif chez la femme	• Codes CIM-10 utilisés (ALD et PMSI [DP ou DR]) : **C50 et D05**
Cancer colorectal actif	• Codes CIM-10 PMSI [DP ou DR]) : **C18- C20, D01.0 - D01.2** • Codes CIM-10 ALD : **C18- C20**
Cancer du poumon actif	• Codes CIM-10 PMSI [DP ou DR]) : **C33- C34, D02.1 - D02.2** • Codes CIM-10 ALD : **C33- C34**
Cancer de la prostate actif	• Codes CIM-10 PMSI [DP ou DR]) : **C61 , D07.5** • Codes CIM-10 ALD : **C61** • Codes ATC utilisés : **G03HA01, L01CD04, L01XX11, L02AA01, L02AA04, L02AE01, L02AE02, L02AE03, L02AE04, L02AE05, L02BB01, L02BB02, L02BB03, L02BB04, L02BX02, L02BX03, V10BX01, V10XX**

Gold standard : Plateforme de Validation des Diagnostics

1

Sélection des cas

SNDS + QM

2

Enquête téléphonique

3

Collecte documents

Anapath, CR hospit.
imagerie, courriers

4

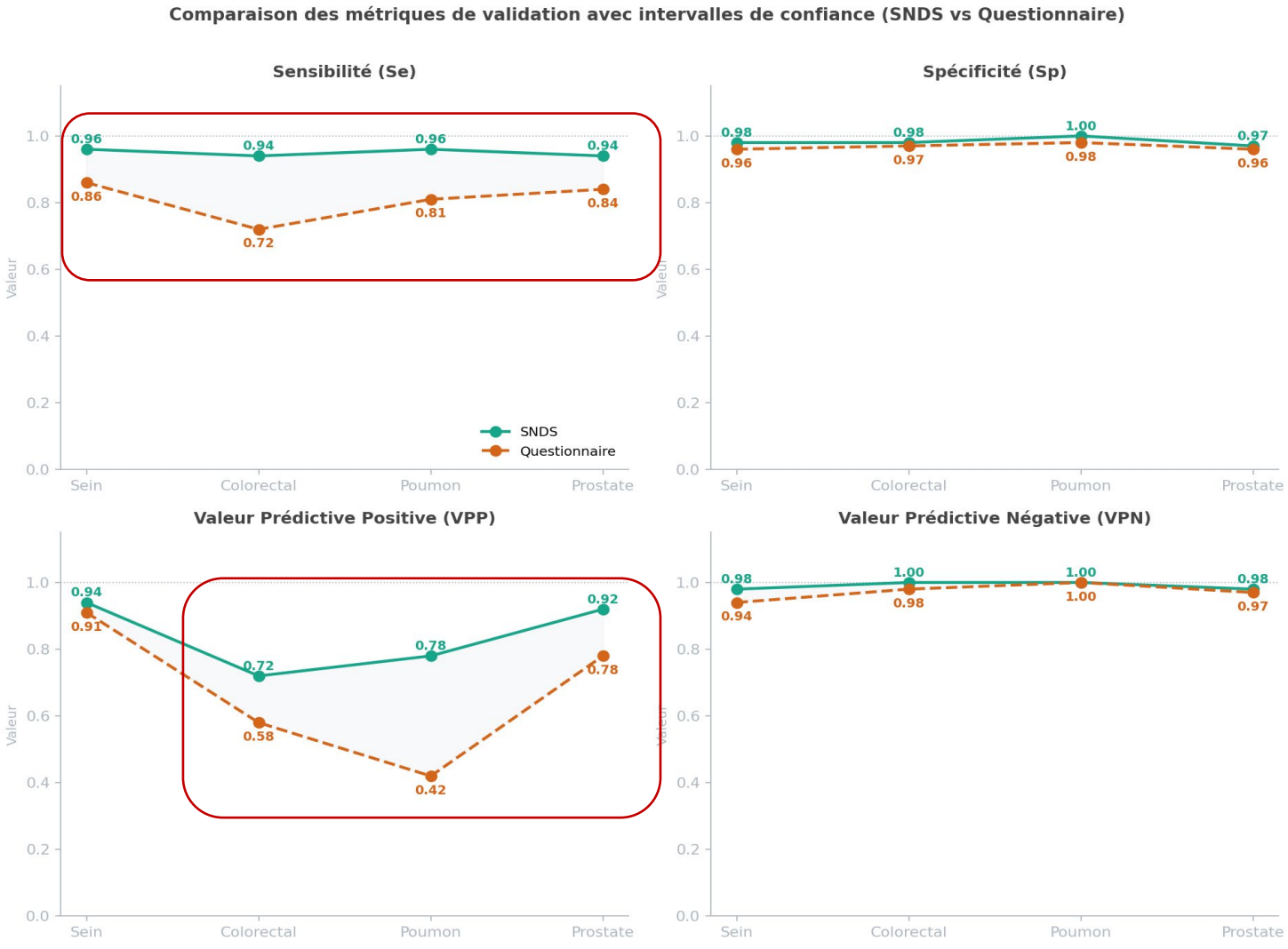
Adjudication FRANCIM

2 experts/dossier
Superviseur si désaccord

4 041	79,9 %	3 411	8 lots
Enquêtes totales	Avec entretien téléphonique	Localisations adjudiquées	D'enquêtes 2014–2024

Informations codées : Date d'incidence · CIM-10 · CIM-O3 (topographie, morphologie) · Grade histologique · Marqueurs pronostiques (PSA, Gleason, RO/RP, HER2, Ki67, Breslow)

Performances de validation



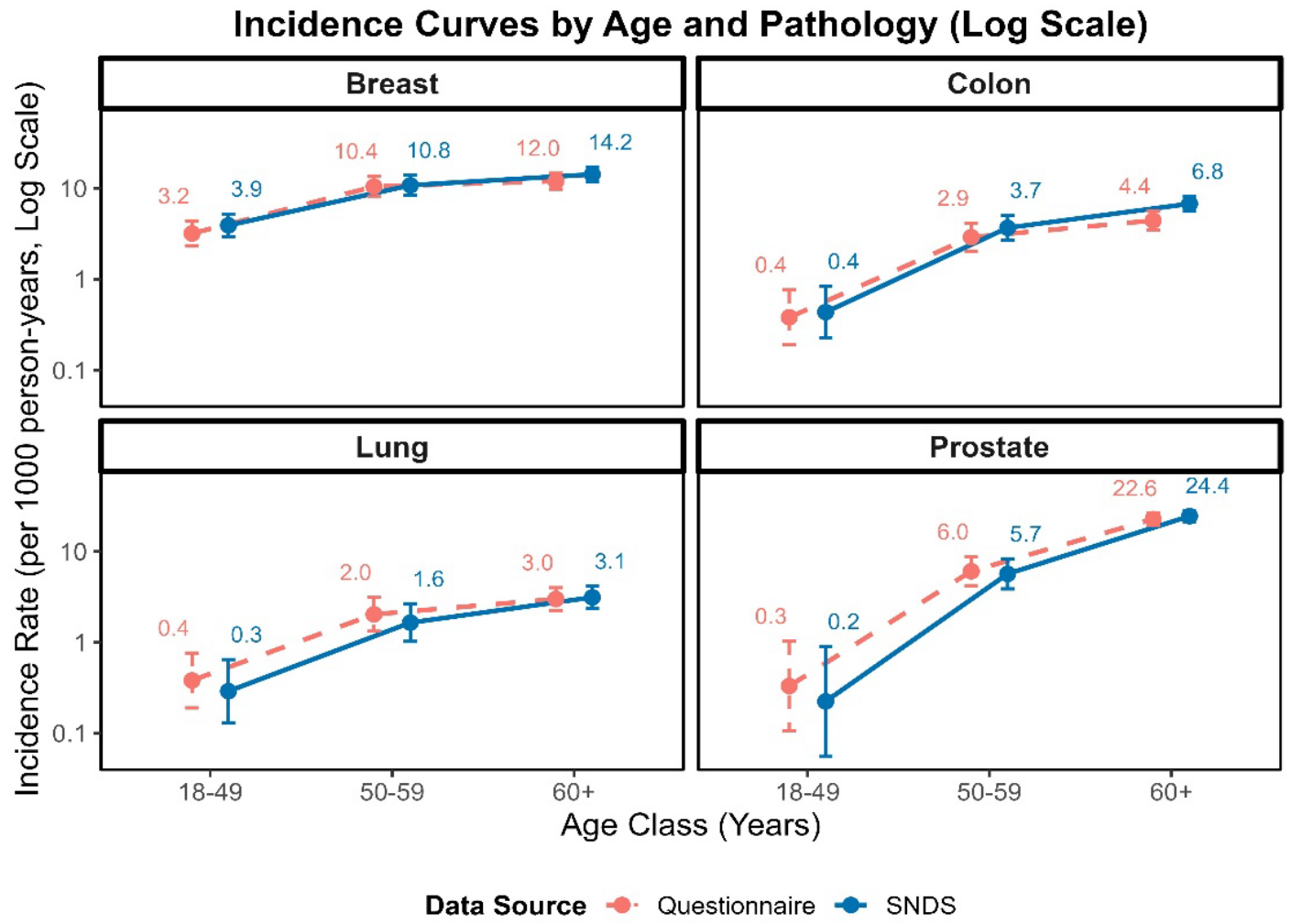
Exploration des cas discordants

Source	CANCER DU POUMON	CANCER COLORECTAL	CANCER PROSTATE	CANCER SEIN
SNDS	Confusion de localisation	Problème de gravité	Confusion diagnostique	Confusion stade
	Métastases (venant du rein, sein, prostate) ou des **cancers guéris** (prévalents) réactivés par le suivi.	L'algorithme capture des **actes de prévention** (retrait de polypes, adénomes) ou de **surveillance** comme s'il s'agissait de traitements de cancer.	**Prend la biopsie** (recherche) pour un traitement, même négative.	Prend le **cancer in situ** (D05) pour un cancer invasif (traitements similaires).
QM	Confusion de localisation ou définition	Confusion définition	Confusion définition	Confusion définition
	Métastases (venant du rein, sein, prostate) ou ou déclare des maladies graves (Amiante, Sarcoïdose) comme cancer.	Le patient assimile **toute chirurgie** ou douleur digestive (Diverticulite, Kystes, inflammations) à un cancer.	Le patient confond l'Hypertrophie (vieillissement) avec le Cancer.	La patiente assimile kyste ou fibroadénome à un cancer.

03.

Incidence des 4 cancers dans Constances

Incidence dans Constances



04.

Constances : Alcool - AUDIT via questionnaires

111. ***Habituellement,*** *à quelle fréquence consommez-vous des boissons alcoolisées ?*

- ☐ 1 1 à plusieurs fois par semaine
  - ➥ Indiquez combien de jours par semaine : ☐☐ jour(s)
- ☐ 2 2 à 3 fois par mois
- ☐ 3 1 fois par mois ou moins
- ☐ 4 Jamais
  - ➥ Indiquez pourquoi :
    - ☐ 1 Pour des raisons de santé
    - ☐ 2 Par goût
    - ☐ 3 Autre

Les 3 questions suivantes portent sur votre consommation de boissons alcoolisées standard.
Pour y répondre, référez-vous aux illustrations ci-dessous.

Une boisson alcoolisée standard

25 cl de bière à 5°	12,5 cl de champagne à 12°	12,5 cl de vin à 12°	2,5 cl de pastis à 45°	4 cl de digestif à 40°	4 cl de whisky à 40°	2,5 cl de digestif à 45°

Tous ces verres standards contiennent la même quantité d'alcool (10 grammes)
Une « chope » de bière (de 50 cl) ou une double « dose » d'alcool fort équivalent à deux boissons standard.

112. ***Les jours où vous consommez de l'alcool,*** *combien de boissons alcoolisées standard buvez-vous* ***en moyenne au cours d'une journée ?***

☐☐ boisson(s) alcoolisée(s) standard

113. *A quelle fréquence vous arrive-t-il de boire* ***six boissons alcoolisées standard ou plus*** *au* ***cours d'une même occasion ?***

- ☐ 1 Jamais
- ☐ 2 Moins d'une fois par mois
- ☐ 3 Chaque mois
- ☐ 4 Chaque semaine
- ☐ 5 Chaque jour ou presque

114. *Au cours des* ***12 derniers mois,*** *quel est le nombre* ***maximal*** *de boissons alcoolisées standard que vous avez consommées en* ***1 jour ?***

☐☐ boisson(s) alcoolisée(s) standard

115. *En pensant aux* ***12 derniers mois,*** *répondez aux questions suivantes :*

	Jamais	Moins d'une fois par mois	Chaque mois	Chaque semaine	Chaque jour ou presque
Combien de fois n'avez-vous pas pu faire ce que normalement vous auriez dû faire, parce que vous aviez bu ?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Combien de fois avez-vous observé que vous n'étiez plus capable de vous arrêter de boire après avoir commencé ?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Combien de fois avez-vous eu un sentiment de culpabilité ou de regret après avoir bu ?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Combien de fois avez-vous été incapable de vous souvenir de ce qui s'était passé la veille parce que vous aviez bu ?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5
Après une période de forte consommation, avez-vous dû boire de l'alcool dès le matin pour vous remettre en forme ?	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5

116. *Au cours des* ***12 derniers mois,*** *vous êtes-vous blessé ou avez-vous blessé quelqu'un parce que vous aviez bu ?*

☐ 1 Oui ☐ 2 Non

➥ **Si non,** cela vous est-il arrivé **auparavant** ? ☐ 1 Oui ☐ 2 Non

117. *Au cours des* ***12 derniers mois,*** *est-ce qu'un parent, un ami, un médecin ou un autre professionnel de la santé s'est déjà préoccupé de votre consommation d'alcool ou vous a conseillé de la diminuer ?*

☐ 1 Oui ☐ 2 Non

➥ **Si non,** cela vous est-il arrivé **auparavant** ? ☐ 1 Oui ☐ 2 Non

Constances : Tabac et obésité via questionnaires/CES

Tabac

104. *Au cours de **votre vie,** avez-vous déjà consommé du tabac (au moins 100 cigarettes (soit 5 paquets) ou 50 cigarillos ou 50 pipes ou 25 cigares) ?*

☐ 1 Oui ☐ 2 Non

➥ **Si non,** passez **directement** à la **question 107, page 27**.

➥ **Si oui** :

- A quel âge avez-vous commencé ? ans
- Fumez-vous encore **actuellement** ? ☐ 1 Oui ☐ 2 Non

➥ **Si non,** à quel âge avez-vous arrêté **(dernier arrêt)** ? ans

IMC

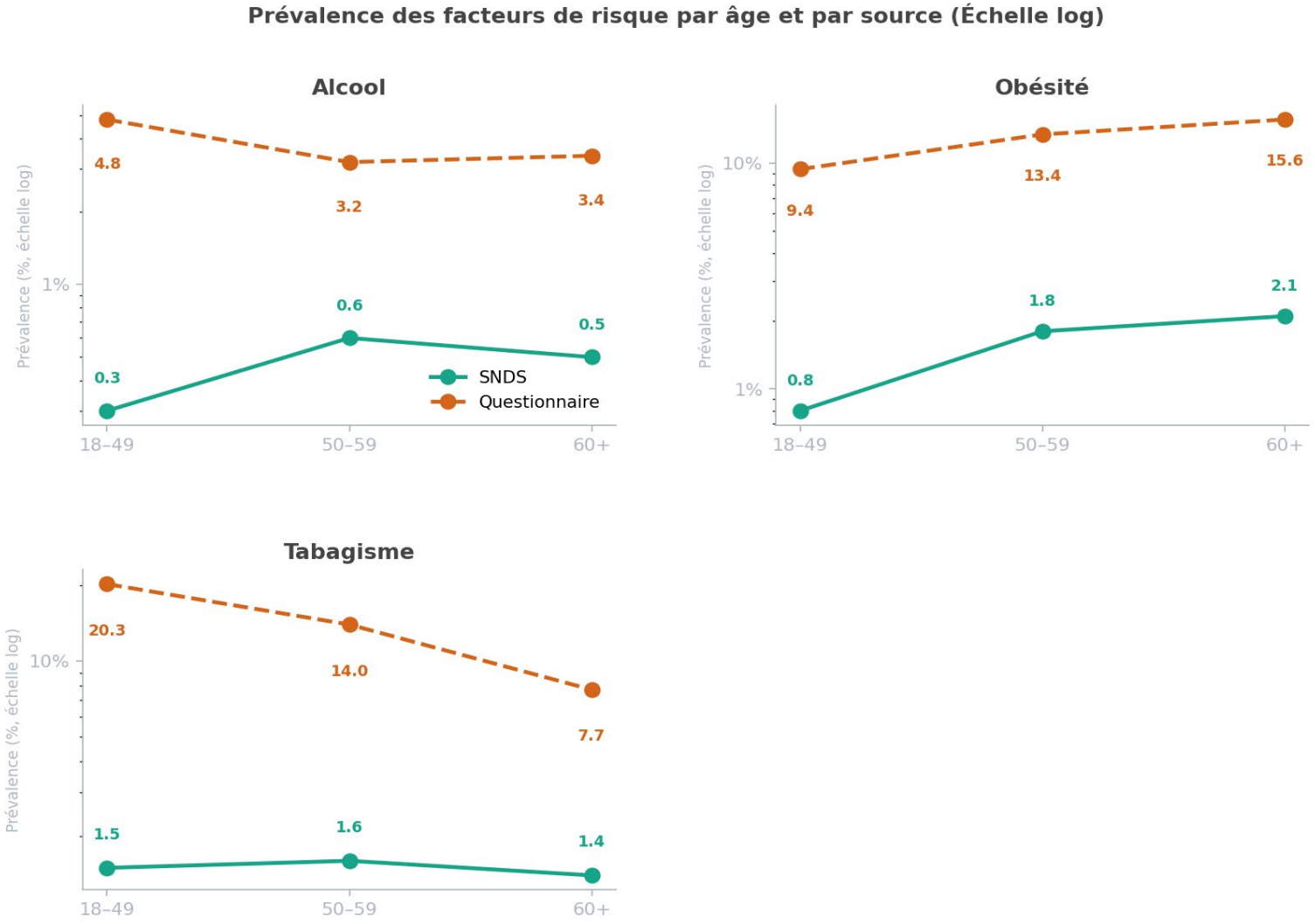
Le bilan Constances est standardisé, chaque examen est réalisé systématiquement **selon un protocole opératoire standardisé** (voir protocoles ci-contre), il s'agit des examens suivants :

- Mesures biométriques (poids, taille, tour de taille, tour de hanches)

Constances : Alcool, Tabac, obésité via SNDS (EPIPHAR)

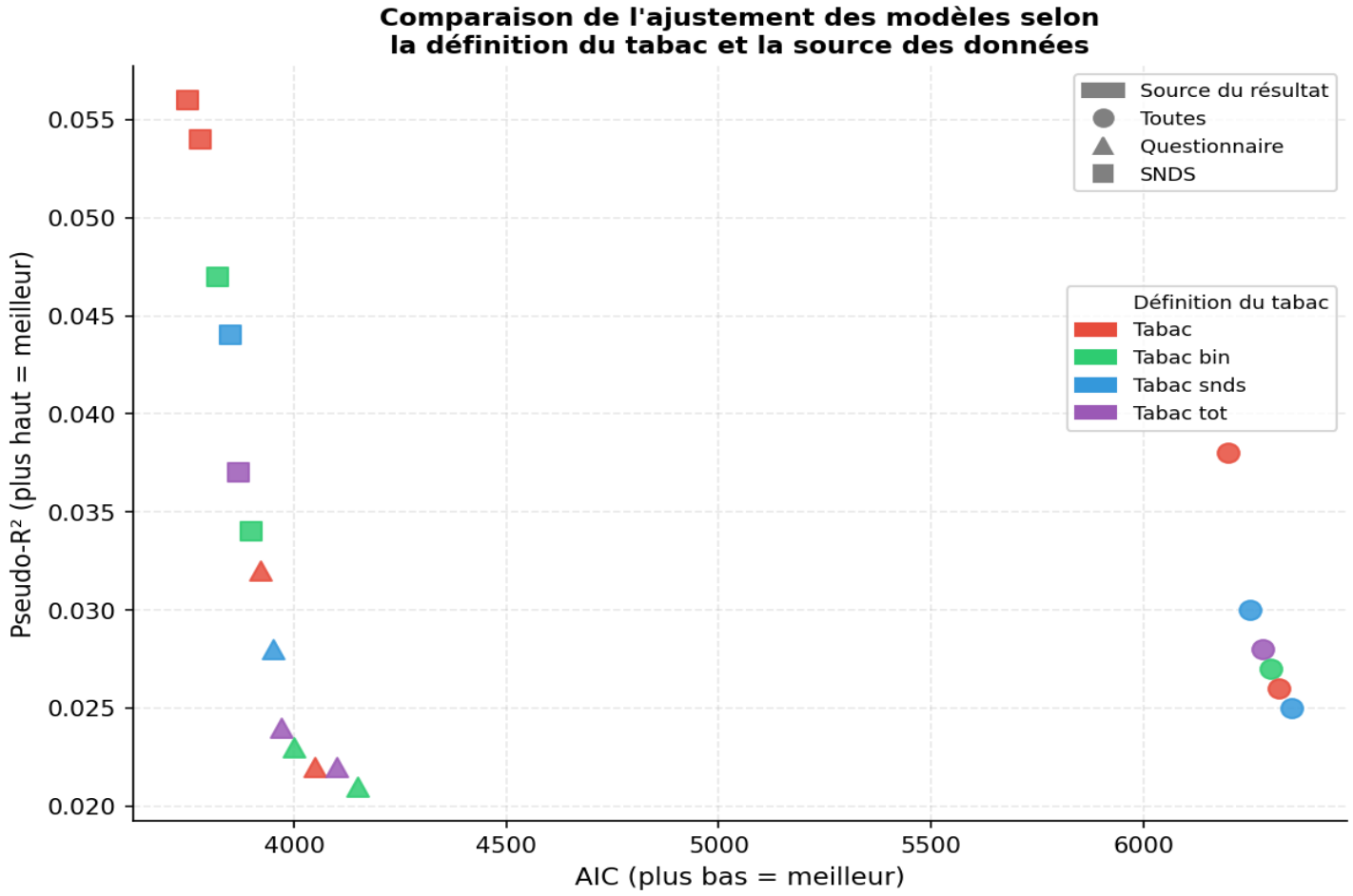
Exposition	**Identification**
Alcool	Présence d'au moins un des critères suivants dans les 5 ans avant l'inclusion • **≥ 1 hospitalisation** (CIM-10 : E244, F10, G31.2, G62.1, G72.1, I42.6, K29.2, K70, K86.0, Z50.2, Z71.4, Z72.1) • **ALD** F10 ou K70 • **≥ 3 médicaments** : disulfirame (N07BB01), acamprosate (N07BB03), naltrexone (N07BB04), nalméfène (N07BB05)
Tabac	Présence d'au moins un des critères suivants dans les 5 ans avant l'inclusion • **≥ 1 hospitalisation** (CIM-10 : F17, Z71.6, Z72.0) • **ALD** F17 • **≥ 3 médicaments** : nicotine (N07BA01), varénicline (N07BA03), bupropion (N06AX12) • **≥ 1 prestation** de prise en charge du tabac (codes 9526, 9527, 9566)
Obésité	Présence d'au moins un des critères suivants dans les 5 ans avant l'inclusion • **≥ 1 hospitalisation** (CIM-10 : E66, T85.50) • **≥ 1 acte chirurgical bariatrique** (CCAM) : bypass gastrique, sleeve gastrectomy, gastroplastie par anneau ajustable, ballon intragastrique, court-circuit biliopancréatique, ou acte de repositionnement/ablation d'anneau

Prévalence des facteurs de risque en fonction de la source

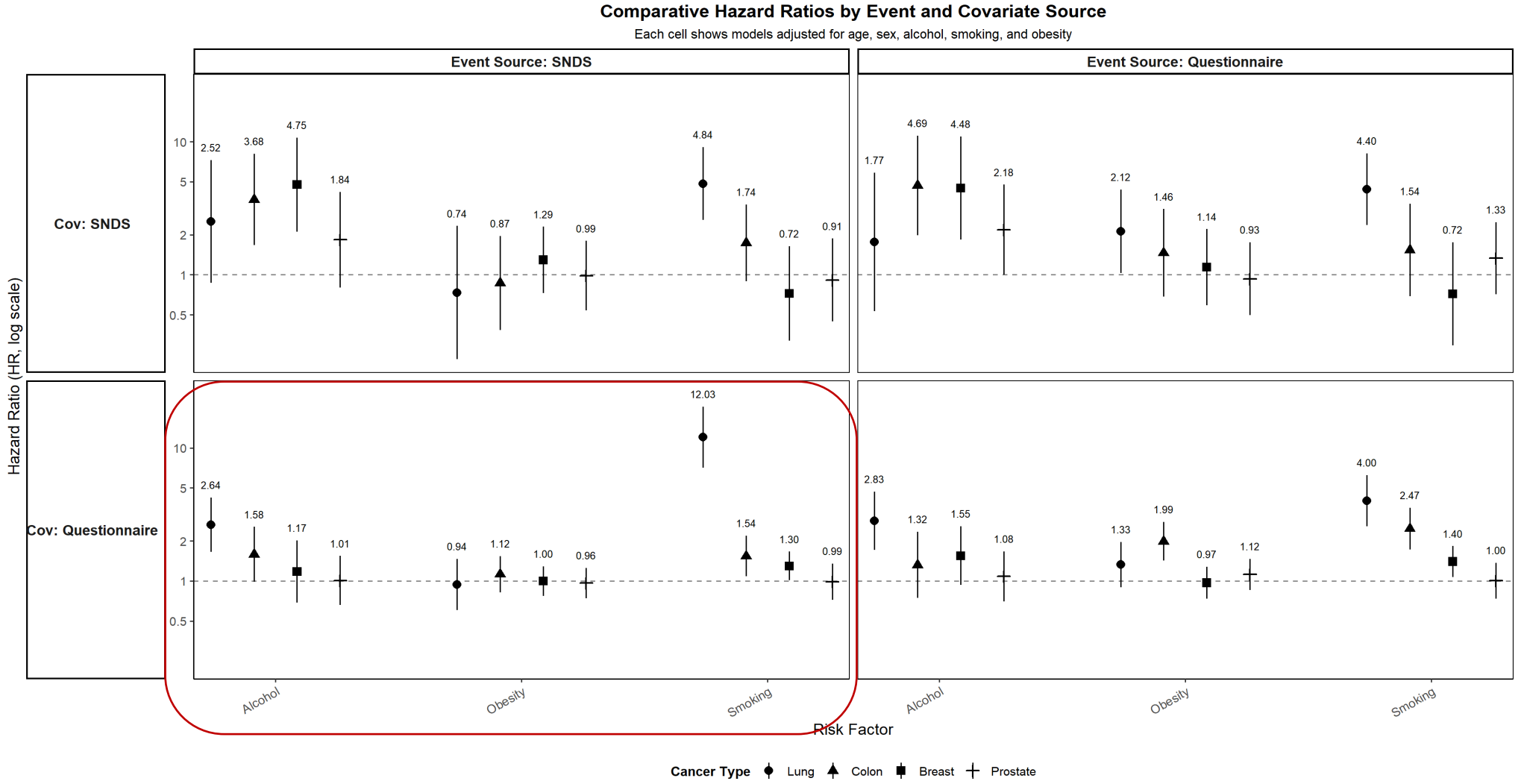


Qualité du modèle en fonction de la combinaison des sources

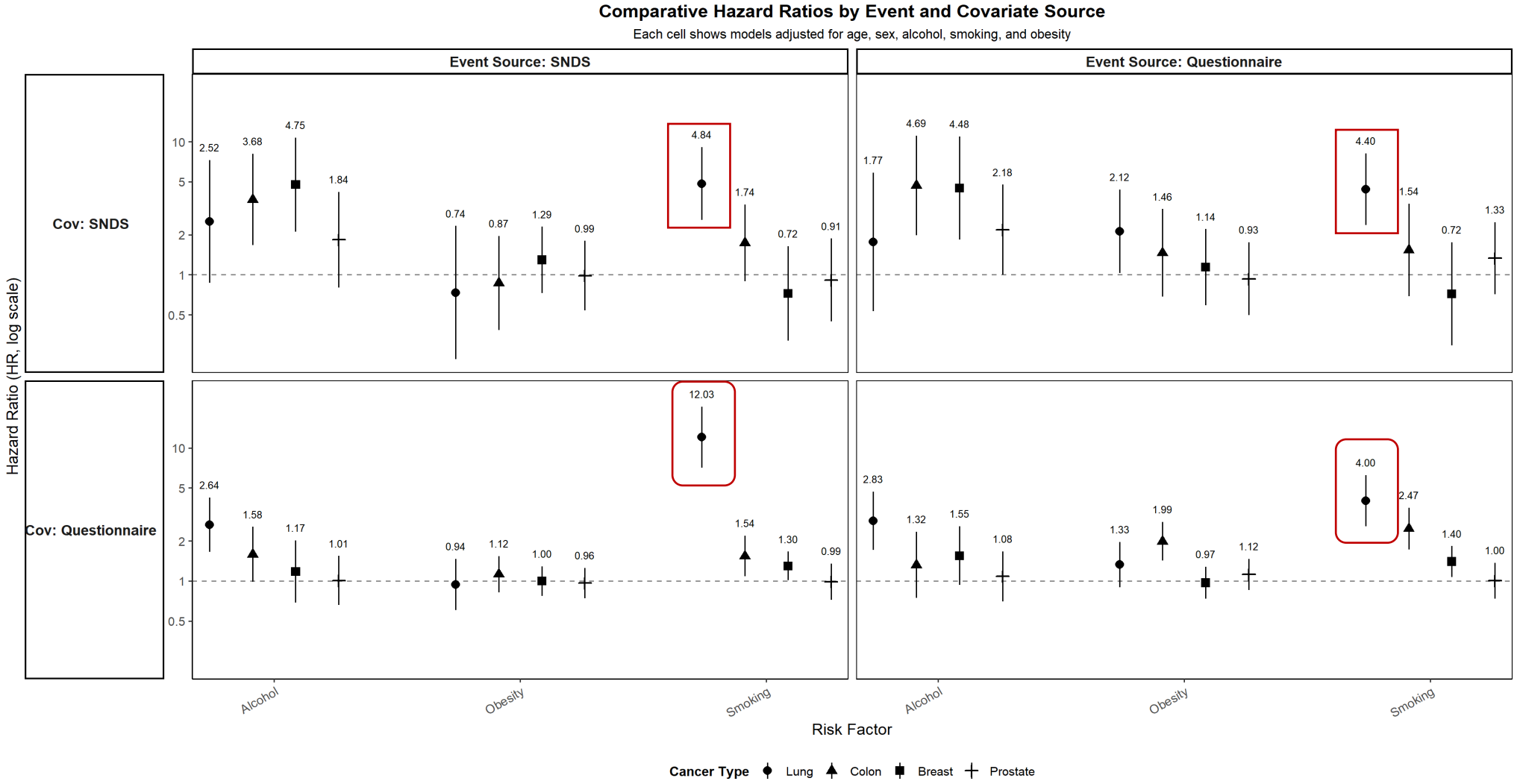
Cancer du poumon — Chaque point = une combinaison exposition × résultat



Association entre cancers et facteurs de risques en fonction de la source



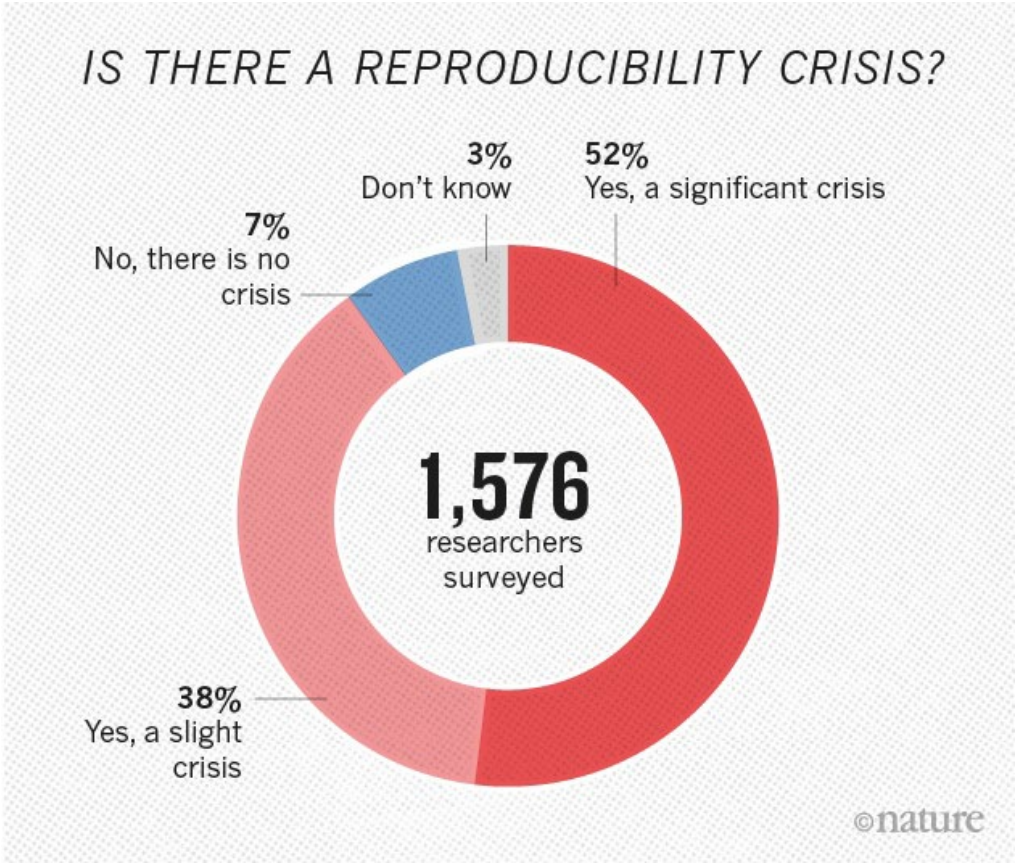
Association entre cancers et facteurs de risques en fonction de la source



05.

Conclusion générale

Contexte for débat sur la validité/reproductibilié des résultats



Why Most Published Research Findings Are False

John P. A. Ioannidis

Published: August 30, 2005 • https://doi.org/10.1371/journal.pmed.0020124

Article | Authors | Metrics | Comments | Media Coverage

Correction
Abstract
Modeling the Framework for False Positive Findings
Bias
Testing by Several Independent Teams
Corollaries
Most Research Findings Are False for Most Research Designs and for Most Fields

Correction

25 Aug 2022: Ioannidis JPA (2022) Correction: Why Most Published Research Findings Are False. PLOS Medicine 19(8): e1004085. https://doi.org/10.1371/journal.pmed.1004085 | **View correction**

Abstract

Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of other studies on the same question, and, importantly, the ratio of true to no relationships among the relationships probed in each scientific field. In this framework, a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are

La source de données n'est pas neutre - façonne les résultats

Sources multiples de bonne qualité -> validité de la mesure -> validité des résultats

- Différentes sources capturent des dimensions complémentaires de l'objectif étudié
- Le croisement des données améliore la validité de l'identification des cas et des mesures d'expositions

Merci !