

# Cécile Sommen

PHD - DATA SCIENTIST / BIOSTATISTICIENNE

Mines Paris PSL - ISPED

685845559 | cecile.sommen@proton.me | 0009-0005-4652-0575 | cecilesommen | cécile-sommen-101a70167

## Profil

Diplômée en Data Sciences par les Mines Paris-PSL et en Biostatistiques par l'Isped (PhD) avec 20 ans d'expérience.  
Spécialiste en biostatistiques, industrialisation des traitements et structuration de pipelines automatisés.  
Forte valeur ajoutée en pilotage transversal, transfert de compétences, production d'indicateurs stratégiques.  
Collaborations internationales en Europe et en Asie, avec des structures de natures diverses.

## Expérience

### Institut Pasteur du Cambodge

Phnom Penh, Cambodge

DATA SCIENTIST

2024 - présent

- Traitements, analyses et valorisation de données provenant de plans de sondage complexes (GLM, GLMM, modèles de survie).
- Appui méthodologique auprès des équipes locales dans un contexte multiculturel et transfert de compétences des bonnes pratiques de programmation et en méthodologie statistique sous R.
- Collaborations scientifiques internationales avec experts, doctorants et post-doctorants, conduisant à des publications internationales.

### Projet Mines Paris - PSL

Saint Maurice, France

DATA-SCIENTIST : CLASSIFICATION DE PRODUITS ISSUS DE LA PLATEFORME RAKUTEN À PARTIR DE DONNÉES TEXTE ET IMAGE

2022

(PYTHON)

- Implication forte dans l'organisation du travail d'équipe (3 personnes) : définition des objectifs, répartition des tâches et suivi du projet.
- Conception et développement d'un pipeline complet de classification multimodale avec structuration des traitements et optimisation des performances et des temps de calcul.
- Gestion de volumes massifs de données dans un environnement cloud (SSPCloud contenant Spark).
- Intégration de données texte et image. Implémentation de réseaux de neurones profonds avec transfert learning et NLP.

### Santé publique France - Direction DATA

Saint Maurice, France

ACTIVITÉS DATA ENGINEER / SCIENTIST

2012 - 2023

- Conception et industrialisation de pipelines de production d'indicateurs nationaux. Automatisation complète des traitements via R et GitLab. Documentation des traitements, des choix méthodologiques et des pipelines.
- Déploiement d'une infrastructure d'orchestration avec Apache Airflow (exécutions parallèles, gestion d'erreurs, alertes, diffusion automatisée).
- Intégration et fusion de bases volumineuses (SIVIC, VAC-SI, SIDEPI) et de bases externes (via API).
- Visualisation et communication : dashboards RShiny (Euromomo), rapports et présentations RMarkdown/Quarto interactifs.

#### ACTIVITÉS BIOSTATISTIQUES ET MODÉLISATION - ÉCHANTILLON DE PROJETS

- Modélisation (GAM) en temps réel de la morbi-mortalité attribuables à la grippe (données multisources : réseau Sentinelles, Renal, PMSI, météo).
- Gestion des données de DO VIH (exhaustivité, imputation des données manquantes, incidence, analyses des facteurs de risque).
- Mise en place de plans de sondage et analyses d'enquêtes (PrevaGay, Coquelicot), estimation de l'incidence et prévalence des IST.
- Estimation de l'efficacité vaccinale de la Grippe et de la Covid19.

#### AUTRES ACTIVITÉS

- Valorisation des travaux (30 publications internationales) et représentation de SpF au niveau national et international (DGS, ANS, ANRS, ECDC...).
- Encadrement de thèse (C. Castel, 2022) et de stages de M2, transfert de compétences aux collègues (formations, guides techniques).
- Participation à l'organisation de séminaires (Rencontres de Santé publique France, Séminaire des Modélisateurs, Séminaire Réseau Sentinelles).

### Observatoire Régional de Santé IDF

Paris, France

BIOSTATISTICIENNE

2009 - 2011

### Danone

Paris-Saclay, France

CONSULTANTE EN BIOSTATISTIQUES

2008 - 2009

### ISPED

Bordeaux, France

MONITEURE DE L'ENSEIGNEMENT SUPÉRIEUR

2006 - 2008

# **Formation**

---

## **Habilitation à l'utilisation des données du SNDS**

HEALTH DATA HUB

2023

## **Diplôme de Data-scientist**

LES MINES PARIS-PSL - 400 HEURES DE FORMATION

2022

## **Thèse en biostatistiques ISPED**

MODÈLES POUR L'ESTIMATION DE L'INCIDENCE DE L'INFECTION PAR LE VIH EN FRANCE À PARTIR DES DONNÉES DE SURVEILLANCE VIH ET SIDA - DIRECTEUR : PR ALIOUM - RAPPORTEURS : PR DE ANGELIS, PR COSTAGLIOLA

2006 - 2009

## **Master 2 Biostatistiques**

UNIVERSITÉ DES SCIENCES DE MONTPELLIER / UNIVERSITÉ DE MONTRÉAL

2004 - 2006

## **Licence MASS option informatique**

UNIVERSITÉ DES SCIENCES DE MONTPELLIER

2001 - 2004

# **Compétences**

---

## **Langues**

ANGLAIS COURANT (UTILISATION PROFESSIONNELLE ET PERSONNELLE QUOTIDIENNE DEPUIS AOÛT 2023)

## **Data Engineering**

BASH, CSS, HTML, YAML, JSON, AIRFLOW, GIT (GITLAB, GITHUB), SPARK, DRAW.IO

## **Statistiques et Data Science**

R, PYTHON, SAS, STATA, LATEX, QUARTO, RMarkdown, RSHINY, STREAMLIT

## **Environnements et éditeurs**

WINDOWS, LINUX, MAC, RSTUDIO, VISUAL STUDIO CODE, JUPYTER

## **Microsoft Office**

EXCEL, POWERBI, POWERPOINT, WORD, ONENOTE, OUTLOOK

# **Projets personnels et bénévolat**

---

## **Développement d'une application de gestion d'une bibliothèque**

WEB API, JSON PARSING, R

2025

- Application Shiny permettant de répertorier rapidement une grande masse de livres en recherchant des livres via ISBN ou titre/auteur, d'éditer leurs métadonnées, puis de les enregistrer dans un fichier Excel structuré.
- [https://github.com/CecileSommen/Book\\_Management\\_AppShiny](https://github.com/CecileSommen/Book_Management_AppShiny)

## **Développement d'une application de vente pour l'association La Petite Librairie Solidaire**

APPSHEET, GOOGLE SHEET

2024

- Application permettant de faciliter la gestion comptable. Outil permettant également à chaque vendeur de saisir les informations de vente pour chaque client, de calculer instantanément le montant total, d'enregistrer automatiquement les données dans un Google Sheet et de visualiser des indicateurs agrégés (nombre de clients, montant des ventes et livres vendus).

## **Développement d'un programme réservation automatique de terrain de tennis**

WEBSRAPPPING, PYTHON, CRON

2021

- Programme permettant de réserver automatiquement un terrain de tennis dès l'ouverture des réservations.
- [https://github.com/CecileSommen/Automate\\_resa\\_tennis](https://github.com/CecileSommen/Automate_resa_tennis)

# Publications

---

Tebben, K., Eng, V., **Sommen C.**, et al. (2025), Parasite clearance in patients with Plasmodium vivax monoinfection treated with artesunate in Cambodia: an observational secondary analysis of trial data, *The Lancet Microbe*.

Eng, V., Lek, D., **Sommen C.**, et al. (2025), 14 days of high-dose versus low-dose primaquine treatment in patients with Plasmodium vivax infection in Cambodia: a randomised, single-centre, open-label efficacy study, *The Lancet Infectious Disease*

Tan-Lhernould, L., Tamandjou, C., **Sommen C.**, et al. (2023), Impact of vaccination against severe COVID-19 in the French population aged 50 years and above: a retrospective population-based study, *BMC Medicine*

Bernadou, A., **Sommen C.**, Pivette M., et al. (2023), Estimating the burden of influenza-attributable severe acute respiratory infections on the hospital system in Metropolitan France, 2012-2018, *BMC Infectious Diseases*

Bonaldi C., Fouillet A., **Sommen C.**, et al. (2023), Monitoring the reproduction number of COVID-19 in France: estimates compared from 3 datasets, *Plos one*

Pillonel J., Maugard C., **Sommen C.**, et al. (2022), Risk of a blood donation contaminated with hepatitis E virus entering the blood supply before the implementation of universal RNA screening in France, *Vox Sanguinis*

Vaux S., Blondel C., [...], **Sommen C.**, et al. (2022), Vaccine coverage against COVID-19 and impact on the dynamics of the epidemic, *Bull Epidémiol Hebd.*

Haneef R., Fayad M., **Sommen C.**, et al. (2022), Direct impact of COVID-19 by estimating disability-adjusted life years at national level in France in 2020, *Plos one*

Cazein F., Bruyand M., [...], **Sommen C.**, et al. (2021), HIV diagnoses in transgender people, France 2012-2020, *Bull Epidémiol Hebd.*

Castel C., **Sommen C.**, Chatignoux E., et al. (2021), Bayesian multi-response nonlinear mixed-effect model: application of two recent HIV infection biomarkers, *The International Journal Of Biostatistics*

Castel C., **Sommen C.**, Le Strat Y., et al. (2021), A multi-state Markov model using notification data to estimate HIV incidence, number of undiagnosed individuals living with HIV, and delay between infection and diagnosis: illustration in France, 2008-2018, *Statistical Methods in Medical Research*

Pantazis N., Rosinska M., [...], **Sommen C.**, et al. (2021), Discriminating between pre-migration and post-migration HIV acquisition using surveillance data, *Journal of Acquired Immune Deficiency Syndromes*

Spaccaferri G., Larrieu S., [...], **Sommen C.**, et al. (2020), Early assessment of the impact of mitigation measures to control COVID-19 in 22 French metropolitan areas, October to November 2020, *Eurosurveillance*

Trouiller P., Velter A., [...], **Sommen C.**, et al. (2020), Injecting drug use during sex (known as “slamming”) among men who have sex with men: Results from a time-location sampling survey conducted in five cities, France, *International Journal of Drug Policy*

Stefic K., Guinard J., [...], **Sommen C.**, et al. (2019), Screening for Human Immunodeficiency Virus infection using a fourth-generation antigen/antibody assay and dried blood spots: in depth analysis of sensitivity and performance assessment in a cross-sectional study, *Journal of Clinical Microbiology*

Brouard C., Saboni L., [...], **Sommen C.**, et al. (2019), HCV and HBV prevalence based on home blood self-sampling and screening history in the general population in 2016: contribution to the new French screening strategy Article, *BMC Infectious Diseases*

Tran T., Pillonel J., [...], **Sommen C.**, et al. (2019), Antenatal HIV screening: results from the National Perinatal Survey, France, 2016, *Eurosurveillance*

Vaux S., Chevaliez S., [...], **Sommen C.**, et al. (2019), Prevalence of hepatitis C infection, screening and associated factors among men who have sex with men attending gay venues: a cross-sectional survey (PREVAGAY), France, 2015., *BMC Infectious Disease*

Ingelbeen B., Bruyand M., [...], **Sommen C.**, et al. (2018), Emerging Shiga-toxin-producing Escherichia coli serogroup O80 associated hemolytic and uremic syndrome in France, 2013-2016: differences with other serogroups., *PlosOne*

Sauvage C., Sabonie L., **Sommen C.**, et al. (2018), Connaissance et utilisation de la prophylaxie pré-exposition (PrEP) parmi les hommes ayant des relations sexuelles avec des hommes fréquentant les lieux de convivialité gay de cinq villes françaises. Prevagay 2015., *Bulletin Epidémiologique Hebdomadaire*

**Sommen C.**, Saboni L., et al. (2018), Time location sampling in men who have sex with men in the HIV context: the importance of taking into account sampling weights and frequency of venue attendance., *Epidemiology and Infection*

Vaux S., Chevaliez S., [...], **Sommen C.**, et al. (2018), Prévalence de l'infection par le virus de l'hépatite B (VHB) et couverture vaccinale contre le VHB chez les hommes ayant des relations sexuelles avec des hommes fréquentant des lieux de convivialité gay de cinq villes françaises. Étude PREVAGAY 2015., *Bulletin Epidémiologique Hebdomadaire*

Van Cauteren D., Le Strat Y., **Sommen C.**, et al. (2018) , Estimation de la morbidité et de la mortalité liées aux infections d'origine alimentaire en France métropolitaine, 2008-2013., *Bulletin Epidémiologique Hebdomadaire*

Noel A., Ruella M., [...], **Sommen C.**, et al. (2017), Large outbreak of urogenital schistosomiasis acquired in Southern Corsica: monitoring the early signs of the endemicization?, *Clinical microbiology and infection*

Trouiller Gerfaux P., Saboni L., [...], **Sommen C.**, et al. (2017), 6 Rapports Prevagay 2015,[ Lille, Montpellier, Lyon, Nice, Paris, National]. Enquête de séroprévalence du VIH menée auprès des hommes ayant des relations sexuelles avec des hommes fréquentant les lieux de convivialité gay., *Santé publique France*

Velter A., Sauvage C., [...], **Sommen C.**, et al. (2017), Estimation de la prévalence du VIH chez les hommes ayant des relations sexuelles avec des hommes fréquentant les lieux de convivialité gay de cinq villes françaises PREVAGAY 2015., *Bulletin de Méthodologie Sociologique*

Jauffret-Roustdide M., Serebroshkaya D., **Sommen C.**, et al. (2017), Comparaison des profils, pratiques et situation vis-à-vis de l'hépatite C des usagers de drogues russophones et francophones à Paris. Enquête ANRS-Coquelicot, 2011-2013. Numéro thématique - Hépatites B et C en populations spécifiques., *Bulletin Epidémiologique Hebdomadaire*

Velter A., Saboni L., **Sommen C.**, et al. (2015), Sexual and prevention practices in men who have sex with men in the era of combination HIV prevention: results from the Presse Gays et Lesbiennes survey, France, 2011., *Eurosurveillance*

Velter A., Saboni L., [...], **Sommen C.**, et al. (2015), Echantillons de convenance par Internet et par la presse - Enquête Presse Gays et Lesbiennes 2011.”, *Bulletin de Méthodologie Sociologique*

Van Cauteren D., De Valk H., **Sommen C.**, et al.(2015) , Community Incidence of Campylobacteriosis and Nontyphoidal Salmonellosis, France, 2008–2013.”, *Foodborne Pathogens and Disease*

Pillonel J., Legrand D., **Sommen C.** , et al. (2014), Surveillance épidémiologique des donneurs de sang et risque résiduel de transmission du VIH, de l'HTLV, du VHC et du VHB par transfusion en France entre 2008 et 2010. Numéro thématique. Don de sang : surveillance.”, *Bulletin Epidémiologique Hebdomadaire*

Cazein F., Lot F., [...], **Sommen C.**, et al. (2013), Découvertes de séropositivité VIH et sida France, 2003-2012.”, *Bulletin Epidémiologique Hebdomadaire*

Campese C., Jarraud S., **Sommen C.**, et al. (2013). Legionnaires disease in France: sensitivity of the mandatory notification has improved over the last decade”, *Epidemiology and infection*

Cazein F., Pinget R.,[...], **Sommen C.** , et al. (2013), Découvertes de séropositivité VIH et sida France, 2003-2011.”, *Bulletin Epidémiologique Hebdomadaire*

Beltzer N., Saboni. L,[...], **Sommen C.** , et al.(2012), An 18-year follow-up of HIV knowledge, risk perception, and practices in young adults.”, *Aids*

Grange D., Host S., **Sommen C.** (2012). Les perceptions de la pollution de l'air intérieur en Île-de-France.”, *Observatoire régional de santé*

Lodwick R.K., [...], **Sommen C.**,[...] (2011). HIV in hiding: Methods and data requirements for the estimation of the number of people living with undiagnosed HIV. *Aids*

Beltzer N., Saboni L.,[...], **Sommen C.** (2011). Les connaissances, attitudes, croyances et comportements des Franciliens face au VIH/sida en Ile-de-France en 2010. “, *Observatoire régional de santé*

Pepin P., d'Acremont M.F., [...], **Sommen C.** (2011). Cancer du sein en Ile-de-France : épidémiologie, pratiques du dépistage, orientations pour les actions de promotion de la santé intégrant le dépistage. “, *Observatoire régional de santé*

Gremy I., Grange D., **Sommen C.** (2011). Les perceptions de la pollution de l'air extérieur en Île-de-France.”, *Observatoire régional de santé*

Bex-Capelle V. , Moularat S.,[...], **Sommen C.** (2010). Effets sanitaires des moisissures dans l'habitat. L'enquête Esmha, pilote d'une étude épidémiologique en Île-de-France”, *Observatoire régional de santé*

Gremy I., Grange D., **Sommen C.** (2010). Les perceptions de la pollution de l'air extérieur en Île-de-France.”, *Observatoire régional de santé*

**Sommen C.**, D Commenges, et al. (2010). Estimation of the distribution of infection times using longitudinal serological markers of HIV: implications for the estimation of HIV incidence. *Biometrics*

Johri M., Morales R.E.,[...], **Sommen C.**, et al. (2010). A cross-sectional study of risk factors for HIV among pregnant women in Guatemala City, Guatemala: Lessons for prevention. *Int J of STD AIDS*

**Sommen C.**, A. Alioum, D. Commenges (2009). A multistate approach for estimating the incidence of Human Immunodeficiency Virus by using HIV and AIDS French surveillance data, *Statistics in Medicine*