

Intitulé du stage : Estimation du nombre de décès tenant compte des délais de transmission des données, à l'aide de méthodes de nowcasting.

Référence de l'offre : DATA-STA-2026-03

■ Présentation de Santé publique France

Santé publique France est l'Agence nationale de santé publique française. Etablissement public de l'Etat sous tutelle du ministre chargé de la santé créé par l'ordonnance 2016-246 du 15 avril 2016, elle intervient au service de la santé des populations.

Agence scientifique et d'expertise du champ sanitaire, elle a pour missions :

- 1° L'observation épidémiologique et la surveillance de l'état de santé des populations ;
- 2° La veille sur les risques sanitaires menaçant les populations ;
- 3° La promotion de la santé et la réduction des risques pour la santé ;
- 4° Le développement de la prévention et de l'éducation pour la santé ;
- 5° La préparation et la réponse aux menaces, alertes et crises sanitaires ;
- 6° Le lancement de l'alerte sanitaire.

L'Agence est organisée autour de quatre conseils (Conseil d'Administration, Conseil scientifique, Comité d'Ethique et de Déontologie et Comité d'orientation et de dialogue), de directions scientifiques et transversales, et de directions assurant le support et le soutien à l'activité. L'Agence dispose d'implantations régionales (Cellules régionales) auprès des agences régionales de la santé.

Son programme de travail, arrêté par son Conseil d'administration, s'articule autour de 6 enjeux : Anticipation, préparation et réponse aux menaces de santé publique, dont les épidémies ; Numérique en santé publique ; Santé environnementale, changement climatique et environnement de travail ; Fardeau des maladies et de leurs déterminants, efficacité des interventions et retour sur investissement de la prévention ; Stratégie de prévention, marketing social et approche par populations ; Inégalités sociales, vulnérabilités territoriales.

Son siège est situé à Saint-Maurice (94).

Localisation du stage :

☒ Saint-Maurice

☐ En région, indiquer la région, le département et la ville concernés :

■ Le stage

Contexte :

Description de la Direction : Le stage s'inscrit au sein de la Direction Appui, Traitements et Analyses des données (DATA) de Santé publique France, qui assure un appui transversal à l'ensemble de l'agence pour le traitement, l'analyse et la valorisation des données. Forte d'une cinquantaine d'agents, la direction est structurée en trois unités spécialisées (unité « Applications, big data et surveillance syndromique - ABISS », unité « Appui et méthodes pour les études et investigations dans le domaine de la surveillance - AMETIS », unité « Enquêtes »). Elle intervient sur l'ensemble du cycle de vie des données de santé.

Ses missions couvrent notamment la gestion de données, l'analyse statistique, la géomatique, la métrologie, ainsi que le développement d'outils informatiques d'analyse et de visualisation. Elle pilote ou soutient plusieurs dispositifs structurants, notamment le système de surveillance syndromique [SurSaJUD](#), l'enquête Baromètre santé, le site open-DATA Odissé ainsi que l'exploitation de bases médico-administratives comme le SNDS.

Description du projet : Les modèles de *nowcasting* (littéralement « prévision immédiate ») permettent d'estimer en temps réel des indicateurs épidémiologiques clés — comme des hospitalisations, des passages aux urgences, des décès — en corrigeant les biais liés aux déclarations incomplètes des données récentes¹. Le principe de ces modèles repose sur l'analyse des délais de signalement, soit l'écart entre la survenue d'un événement et son enregistrement dans les bases de données.

Ces modèles exploitent l'historique des bases de données pour identifier, pour chaque événement, sa date réelle et sa date de déclaration. Ces données historiques permettent notamment de construire une distribution des délais de signalement (par exemple, 50 % des cas sont déclarés sous 70h, 95 % sous dix jours, etc.). En croisant ces distributions avec les données observées mais incomplètes, les modèles estiment la proportion d'événements manquants pour les périodes récentes. Par exemple, si seulement 50% des passages aux urgences sont déclarés le jour j , le *nowcasting* extrapole les 50 % restants pour fournir une estimation complète et actualisée du nombre total d'événements. Cependant, en pratique, la situation n'est que rarement aussi simple, la variation des retards de signalement fluctuant selon la période² (sous-déclaration les week-ends ou pendant les fêtes), les lieux³ (selon les territoires ou les établissements) ou le contexte (saturation de la prise en charge entraînant un retard supplémentaire de déclaration). Les modèles de *nowcasting* permettent d'intégrer ces dynamiques pour ajuster les prédictions, ainsi que de quantifier l'incertitude associée aux estimations (intervalles de confiance).

L'objectif du stage sera donc de réaliser une revue des méthodes de *nowcasting*⁴⁻⁹ puis de les mettre en œuvre sur des données réelles afin d'en dégager les forces et les faiblesses dans un objectif de mise en œuvre en routine sur des données de surveillance de l'état de santé de la population.

Données d'application : Le système de surveillance syndromique SurSaUD, piloté par Santé publique France a été mis en place en 2004 à la suite de la canicule de 2003, ce dispositif permet de suivre en temps quasi réel l'état de santé de la population, en facilitant la détection précoce d'événements sanitaires, qu'ils soient attendus ou inhabituels. SurSaUD repose sur l'intégration quotidienne et automatisée de données individuelles et anonymisées issues de quatre sources complémentaires :

- Les passages aux urgences hospitalières (réseau OSCOUR),
- Les interventions des associations SOS Médecins,
- Les certificats électroniques de décès transmis à l'Inserm-CépiDC.
- Les données de mortalité issues des bulletins d'état civil transmis à l'Insee,

Les données de mortalité collectées par les bureaux d'état-civil remontent progressivement à l'Insee et à Santé publique France avec un délai par rapport à la date de décès. En moyenne, 50% des décès survenant un jour J remontent en 3 jours et 95% en 10 jours¹⁰. Ainsi, les effectifs les plus récents sont sous-estimés, ce qui masque les tendances actuelles. Les données sont considérées consolidées au bout de 30 jours. Ce délai peut varier selon le jour de survenu du décès dans la semaine selon qu'il se produit pendant une période de vacances scolaires, un jour férié, en situation d'épidémie, ou encore selon la taille et l'organisation des communes^{2,3}. En pratique, ce délai de transmission rend difficile la surveillance réactive et l'interprétation des évolutions de la mortalité à court terme. Ainsi, chaque début semaine, la surveillance de la mortalité toutes causes analyse les effectifs de la semaine S-2. Il reste donc difficile d'avoir une analyse robuste des effectifs de décès de J-1 à J-10.

L'objectif vise à produire chaque jour J , une estimation rétrospective du nombre de décès survenus entre J-1 et J-30, en corrigeant les biais liés aux retards de déclaration des données en utilisant les méthodes de *nowcasting*⁴⁻⁹. Les modèles seront appliqués sur une période de plusieurs années et couvriront différents niveaux : national, régional et par groupe d'âge. Les performances des modèles seront évaluées à l'aide d'un cadre probabiliste (probabilistic scoring framework), afin d'identifier la meilleure approche. L'enjeu est de proposer un modèle de correction des délais de transmission optimisé pour la surveillance en routine et en temps réel.

Missions du stage : Le stage inclura la participation aux missions suivantes, organisées en cinq axes complémentaires :

1. **Recherche bibliographique des méthodes de nowcasting**
2. **Prise en main et description des données**
 - Description de l'évolution des effectifs de décès, selon le délai de transmission.
 - Déclinaison des évolutions, selon les caractéristiques de période (jour de la semaine, jour férié et pont, période de l'année, zone géographique).
3. **Mise en œuvre et évaluation des performances d'une ou plusieurs méthodes**
 - Nowcasts quotidiens de J-1 à J-30, stratifiés par âge et échelle spatiale pour différentes méthodes de *nowcasting*.
 - Production d'estimations probabilistes (médianes et intervalles de crédibilité).

- Évaluation des performances via un cadre probabiliste, permettant de comparer la qualité prédictive des différentes approches.

4. Tests de sensibilité des méthodes retenues

5. Livraison d'un script pour une utilisation en routine et rédaction éventuelle d'un article scientifique

Environnement de travail : Le stage s'intègre dans un environnement technique dynamique et collaboratif, mobilisant des outils de développement modernes, des langages adaptés à la science des données et des infrastructures de calcul performantes. Le·la stagiaire évoluera au sein d'une équipe pluridisciplinaire, en interaction étroite avec des épidémiologistes, médecins, data scientists, statisticien·ne·s, ingénieur·e·s et membres de la DSI.

Les principaux outils et technologies mobilisés incluent :

- Langages : R, Python
- Environnement collaboratif : GitLab
- Formats et bases de données : csv, txt
- Environnements de développement : R/Python, Rstudio, Jupyter notebook, VS Code, IA Mistral

Profil recherché : Stagiaire de niveau Master 2, issu·e d'une formation en statistique, mathématiques appliquées, physique ou disciplines connexes avec des compétences en data science. Le·la candidat·e devra en particulier maîtriser les langages R et/ou Python.

Le·la candidat·e devra faire preuve de bonnes capacités d'analyse, d'adaptation et d'organisation, ainsi que d'un réel esprit d'équipe pour évoluer au sein d'un environnement pluridisciplinaire.

Références :

1. CDC. Behind the Model: Nowcasting. *CFA: Behind the Model* <https://www.cdc.gov/cfa-behind-the-model/php/data-research/nowcasting.html> (2025).
2. Accounting for Day-of-the-Week Effect in Nowcast Models. https://cran.r-project.org/web/packages/NobBS/vignettes/dow_effect.html.
3. Wolfram, D. *et al.* Collaborative nowcasting of COVID-19 hospitalization incidences in Germany. *PLoS Comput. Biol.* **19**, e1011394 (2023).
4. Lawless, J. F. Adjustments for Reporting Delays and the Prediction of Occurred but Not Reported Events. *Can. J. Stat. Rev. Can. Stat.* **22**, 15–31 (1994).
5. Sam Abbott *et al.* epinowcast: Flexible Hierarchical Nowcasting. <https://doi.org/10.5281/zenodo.5637165> (2025).
6. Schneble, M., De Nicola, G., Kauermann, G. & Berger, U. Nowcasting fatal COVID-19 infections on a regional level in Germany. *Biom. J. Biom. Z.* **63**, 471–489 (2021).
7. van de Kasstelee, J., Eilers, P. H. C. & Wallinga, J. Nowcasting the Number of New Symptomatic Cases During Infectious Disease Outbreaks Using Constrained P-spline Smoothing - PubMed. *Epidemiology* (2019).
8. Fritz, C. *et al.* Statistical modelling of COVID-19 data: Putting Generalised Additive Models to work. *Stat. Model.* **24**, 344–367 (2024).
9. Heyder, S. & Hotz, T. Stochastik-TU-Ilmenau/ILM-prop. Stochastik - Technische Universität Ilmenau (2023).
10. Surveillance syndromique - SURSAUD® – Santé Publique France. <https://www.santepubliquefrance.fr/surveillance-syndromique-sursaud-R>.

Description : Pour tout renseignement complémentaire, consulter le portail sur le site <http://www.santepubliquefrance.fr>

■ Type de stage proposé

☐ Licence

☐ Master 1

☒ Master 2 Professionnel

☐ Master 1 + sujet tutoré

☐ Master 2 Recherche

☐ Master 1 « observation »

Gratification du stage : selon la réglementation en vigueur dans les établissements publics

▪ **Date proposée pour le stage et durée :**

☒ Sans contrainte de date

☐ A partir de : (indiquer une date)

Durée minimum : 6 mois

Extension possible au-delà de la période obligatoire

☐ Oui

☒ Non

▪ **Prérequis :**

☐ Aucun

☐ Compétences spécifiques (préciser) :

☒ Maîtrise d'un logiciel spécifique (préciser) : Langage R, Python

☒ Autre (préciser) : Connaissance sur les outils de versionning de code tels que GitHub/GitLab

▪ **Stage proposé par :**

Direction : DATA

Unité : AMETIS/ABISS

Maître(s) de stage / personne contact :

Nom : Sabbatini

Prénom : Chiara

Fonction : Statisticienne

Téléphone :

Adresse email : chiara.sabbatini@santepubliquefrance.fr

Co-encadrement

Nom : Fouillet

Prénom : Anne

Fonction : Epidémiologiste, Statisticienne

Téléphone :

Adresse email : anne.fouillet@santepubliquefrance.fr

▪ **Pour postuler**

Adresser les candidatures (lettre de motivation + cv) en indiquant la référence de l'annonce par courriel :

recrut@santepubliquefrance.fr