

Appel à projet
Moyens de la recherche 2015-2016
LabEx IRMIA

1-Titre du projet : INCEPTION

Interdisciplinary Network on Causality in EPidemiological investigaTIONS
Réseau interdisciplinaire sur la causalité dans les études épidémiologiques

2-Coordonnées du porteur

Pr. SAULEAU Erik-André

Laboratoire de Biostatistique et Informatique Médicale (responsable)

Faculté de Médecine
4, Rue Kirschleger
F-67085 Strasbourg

Jusqu'au 31/12/2015

Equipe d'Accueil 3430, Progression tumorale et micro-environnement. Approches translationnelles et épidémiologie (Directrice Dr D. Guénot)

A partir du 01/01/2016

Laboratoire iCube, CNRS UMR 7357 (Directeur Pr M. de Mathelin)

Hôpitaux Universitaires de Strasbourg

Pôle de Santé Publique, Groupe Méthode en recherche Clinique
1, Place de l'hôpital
F-67091 Strasbourg Cedex

3-Le projet

Description

Dans les 50 dernières années, les statistiques ont émergé comme la règle pour toute preuve en médecine, fondant la médecine factuelle (EBM, « evidence based medicine »).

Dans la première moitié du 20e siècle, les épidémiologistes ont commencé à décrire la cause de la maladie en termes nouveaux : à partir de la réponse simple d'un agent spécifique et unique (bactérie) vers un concept multifactoriel de causalité [1]. La conséquence même de ce changement est que l'épidémiologiste fait maintenant face à des situations très complexes. Par exemple, le graphique ci-contre montre que plusieurs facteurs peuvent interagir dans la causalité du régime alimentaire sur l'asthme [2]. La légende du graphique montre combien peut être compliquée la prise en compte de tous ces facteurs.

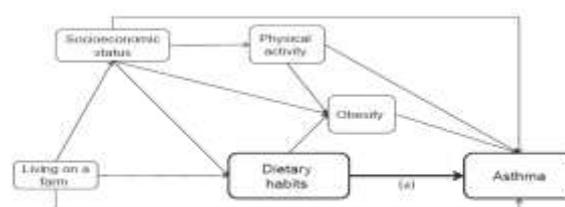


Figure 1 Example of the causal association between diet and asthma, taking into account 'potential confounders' using Direct Acyclic Graph (DAG). Assessing the causal association between dietary habits and asthma (arrow a) needs to take into account strongly correlated 'potential confounders' (1). For example, while a farming environment in childhood may have a protective effect on asthma till adulthood, it might also influence dietary habits till adulthood (6), as well as socioeconomic status later in life. Further, according to this DAG, obesity is a mediator in the effect of diet on asthma, and physical activity is a confounder in the assessment of the effect of obesity on asthma. If we adjust for obesity and physical activity when studying the effect of diet on asthma, we will get biased results.

Basés sur les travaux concernant la randomisation de Ronald A Fisher [3], les essais contrôlés randomisés ont été érigés comme le *gold standard* de la preuve de l'efficacité de tout médicament ou dispositif médical. Mais ces essais ont plusieurs limitations (coûts, éthique, généralisabilité, ...). Par conséquent, de nombreuses données proviennent d'observations non-expérimentales. Ainsi dans les deux cas, un raisonnement sophistiqué sur la causalité est nécessaire. Si la pratique épidémiologique tend à se reposer sur un faisceau

d'arguments étayant la présomption (d'un certain type) de causalité, la littérature (bio)statistique propose plusieurs modèles alors que dans une autre perspective, la littérature philosophique a fourni, et continue de fournir, plusieurs approches de la causalité.

L'**objectif général** du projet est de réfléchir aux fondements philosophiques et épidémiologiques/biostatistiques de la causalité et de la modélisation causale, avec un ancrage tout particulier sur l'historique de ces approches et leur historiographie. L'approche mobilisée est la constitution d'un réseau scientifique international, éminemment multi- et interdisciplinaire. La réponse à l'appel à projet du LabEx ne concerne qu'une partie de la création de ce réseau et plutôt les moyens nécessaires au fonctionnement scientifique du réseau et non ses moyens « logistiques », qui eux feront l'objet d'une demande de cofinancement par l'ANR (dans un premier temps, appel à projet MSREI, Montage de Réseaux Scientifiques Européens ou Internationaux). L'ensemble du projet, intitulé INCEPTION, s'articule autour de trois étapes chronologiques :

1. intensifier et structurer le travail collaboratif des trois équipes actuelles, en favorisant leurs rencontres et leurs capacités de recherche scientifique ;
2. organiser à Strasbourg un colloque fondateur du réseau, qui serait un colloque scientifique mais aussi l'occasion de réunir des équipes de recherche et de les « recruter » ;
3. la dernière étape sera d'assurer la permanence du réseau, tant dans sa logistique que dans ses projets scientifiques.

INCEPTION, dont le Laboratoire de Biostatistique et Informatique Médicale de l'Université de Strasbourg (Faculté de Médecine) sera l'animateur, commence à s'articuler autour d'un axe géographique réunissant trois lieux et une série d'équipes « fondatrices » :

- à Strasbourg, un groupe de travail regroupe plusieurs disciplines et implique actuellement : les Pr. Bonah (Professeur des Universités en histoire de la médecine) et Sauleau (Professeur des Universités - Praticien Hospitalier en biostatistique, Master Science et société - histoire, philosophie et médiation des sciences), M. Schaeffer (Ingénieur en statistique, Master de mathématiques), les Dr Séverac (Assistant Hospitalo-Universitaire, DES de Santé publique, Master d'épidémiologie), qui s'inscrira en thèse de troisième cycle pour la première année (2015-2016) avec un sujet portant sur l'apport des réseaux bayésiens dans l'analyse de la causalité, ainsi que les Dr Lefèvre (Docteur en Médecine, Praticien Hospitalier en épidémiologie), Cuchet et Thiery (DES de Santé publique, thèse d'exercice en cours, Master d'épidémiologie). Cette dernière a montré dans son mémoire de master, en utilisant les modèles structuraux marginaux, que les patients en insuffisance rénale chronique terminale traités par hémodialyse avaient une meilleure survie que ceux sous dialyse péritonéale (deux formes différentes de suppléance de la fonction rénale).
- à Cambridge, le Pr. Dawid (Professeur émérite de l'Université de Cambridge, ancien Directeur du Laboratoire de Statistique de l'Université de Cambridge) positionne ses recherches en mathématique-statistique mais avec une forte composante philosophique ;
- à Cagliari, la Pr. Musio (Professeure des Universités en mathématique) collabore avec le Pr. Dawid et conduit ses recherches en mathématique-statistique.

A ces trois équipes fondatrices du réseau, sera adossée une équipe d'anthropologie des sciences (Dr Allamel-Raffin, Maître de conférences en épistémologie et histoire des sciences et des techniques, Université de Strasbourg) qui observera la constitution du réseau. Elle examinera notamment les échanges au sein du réseau, tant explicites que tacites [4], le rôle des contraintes institutionnelles.

La première étape du projet INCEPTION est donc de conforter ces trois équipes et leur permettre de poursuivre et amplifier leurs travaux scientifiques ainsi que de structurer plus solidement leur collaboration. C'est précisément cette étape qui fait l'objet de cette réponse à l'appel à projet du LabEx. Dans la mesure où un co-financement est demandé à l'ANR pour la constitution du réseau, l'organisation du colloque fondateur sera plus volontiers financée par l'ANR. Mais il va sans dire que le LabEx IRMIA sera identifié comme la structure cadre de l'ensemble du projet INCEPTION.

La réflexion au sein du projet sera étayée par la re-analyse de données d'épidémiologie qui ont déjà été exploitées antérieurement au sein du Pôle de santé publique des hôpitaux universitaires de Strasbourg.

Le colloque fondateur du réseau, organisé à Strasbourg en 2017, sera un colloque scientifique mais aussi l'occasion de réunir des équipes et de « recruter » les futures équipes du réseau. Des contacts préalables

auront été pris, notamment avec le Pr. Schumacher, *Center for Medical Biometry and Medical Informatics*, Université de Freiburg et le Pr. Held, *Epidemiology, Biostatistics and Prevention Institute* de l'Université de Zurich, avec qui des projets de recherche ont déjà été menés. Le colloque fondateur permettra de fixer un jalon (dans le temps et dans l'espace) dans la recherche internationale sur les modèles de causalité. Le colloque d'une semaine sera précédé par une semaine de préparation impliquant les trois équipes. Compte tenu de son caractère multidisciplinaire, un ouvrage pourrait être édité, servant de résumé du colloque.

Enjeux

Ce qui se joue dans le projet est la place occupée par la causalité statistique dans la validation des données épidémiologiques ; l'épidémiologie vue comme un des trois régimes de preuves en médecine (avec la clinique et le laboratoire). Le projet est très interdisciplinaire. Un des enjeux majeur sera donc de coordonner les différentes disciplines. De même la complémentarité des équipes devra être maintenue.

Plus spécifiquement à Strasbourg, l'enjeu est celui du fonctionnement pérenne d'un nouveau groupe de recherche, dans une zone de recherche actuellement non représentée et d'établir une collaboration entre la médecine, la mathématique et les sciences humaines entre les facultés de l'université.

Enfin, INCEPTION est une occasion unique pour l'anthropologie des sciences de voir se créer un réseau et être impliquée dès l'origine (depuis la réflexion bibliographique jusqu'à la mise en action du réseau).

Contexte

Dans la première moitié du 20e siècle, les conceptions scientifiques de causalité ont subi un changement radical [5]. En raison ou en parallèle de la «transition épidémiologique», les études sur les maladies chroniques ont été dirigés vers l'identification des groupes à risque élevé afin de cibler les interventions de santé publique. Les données utilisées proviennent d'essais contrôlés ou d'observations non-expérimentales. Le passage d'expériences aléatoires à des études d'observation exige un examen de « critères de causalité » [6]. Il existe de nombreux exemples d'études épidémiologiques qui utilisent les neuf directives de Bradford Hill (cohérence, force, spécificité, relation temporelle, cohérence, gradient biologique, plausibilité, données expérimentales et analogie) [7] dans une tentative pour justifier les revendications de causalité, mais sans qu'on ait l'assurance de sa réalité [8,9].

En parallèle de cette réflexion des épidémiologistes, plusieurs modèles statistiques causaux ont été développés : modèles graphiques, analyse des chemins et modèles basés sur l'approche contrefactuelle, avec un accent particulier sur les modèles structuraux marginaux [10]. L'approche contrefactuelle définit la causalité en comparant le résultat observé (mourir d'un cancer par exemple) et le résultat contrefactuel : le résultat qui aurait été observé si, contrairement aux faits, le sujet avait reçu une exposition différente (par exemple, n'aura pas été exposé à un facteur de risque donné) que celle qu'il a effectivement reçue.

Une définition moderne de la causalité largement utilisée en philosophie est celle de Hume [11]. Mais il y a un certain nombre de difficultés avec les théories dites de régularité issues de cette acception, qui ont motivé les approches probabilistes à la causalité [12,13]. D'autres modèles (philosophiques) de causalité ont été développés pour représenter les systèmes de relations causales, tels que celui proposé par Pearl [14]. Le but de Pearl [15] est de combler l'écart entre les expressions verbales et la mathématique. Il propose alors des expressions mathématiques pour la confusion, l'analyse contrefactuelle, et plus généralement les principes méthodologiques de l'inférence causale. Un cadre populaire pour les modèles causaux contrefactuel est le modèle préconisé par Lewis [16-17] mais des chercheurs ont proposé d'autres cadres, comme par exemple Dawid [18,19] ou Spirtes, Glymour et Scheines [20]. Pour le moment ces définitions et modèles de causalité n'ont pas connu d'application majeure en épidémiologie.

4-Articulation avec le projet du LabEx IRMIA

Dans les objectifs stratégiques du LabEx figurent le renforcement de la capacité de recherche et d'influence internationale de l'IRMA et l'accroissement des synergies avec les partenaires des mathématiques. L'approche novatrice et originale du projet INCEPTION répond à ces objectifs. Un laboratoire de Strasbourg se charge de coordonner un réseau international scientifique. Des contacts seront d'autre part rapidement pris avec des équipes de Zurich et de Freiburg (mais l'ANR exclut de son financement tout projet qui pourrait être du ressort d'INTERREG).

Le projet de recherche du LabEx est non seulement de maintenir l'excellence en mathématique pure, mais aussi de développer des projets de recherche ambitieux notamment vers les statistiques et leurs interactions avec les autres sciences. L'aspect multidisciplinaire d'INCEPTION s'inscrit dans ce projet. Il est de nature à attirer une nouvelle population d'étudiants dans différentes formations, ce qui est aussi un axe de la politique du LabEx.

5-Le budget

Le budget qui concerne l'intégralité du projet repose sur un financement dans le présent appel d'offre et sur un co-financement de l'ANR dans le cadre de l'appel à projet MRSEI. Cette dernière enveloppe, au maximum de 30k€ sur 18 mois, ne couvre aucun frais de personnels et concerne plutôt les moyens « logistiques » de la création du réseau. On considère que la demande qui sera faite à l'ANR (la date limite de dépôt de la réponse à l'appel d'offre est le 12/01/2016) couvrira les frais d'organisation du colloque de 2017 (pour 25k€) et des frais de rencontres ponctuelles entre les équipes fondatrices du réseau (pour 5k€). Le budget soumis dans le cadre de la présente réponse ne concerne alors qu'une partie de la création du réseau et plutôt les moyens nécessaires au fonctionnement scientifique du réseau. Le budget total demandé au LabEx est de **110 k€ sur 2 ans**. Dans le détail :

- un des enjeux majeur du projet est la réussite du « dialogue multidisciplinaire » et d'autre part, l'essentiel du début du fonctionnement du réseau aura plus volontiers à faire dans les disciplines (bio)statistique, mathématique et épidémiologique. Ce qui s'impose est alors la nécessité d'un ingénieur de recherche, dont l'encadrement serait assuré conjointement Mr Frédéric Bertrand (IRMA) et par le porteur du projet (cf. fiche de poste en annexe). La couverture des projets MRSEI est de 18 mois mais un chevauchement est nécessaire entre les deux présents appels à projet (LabEx et ANR) et l'appel à projet ultérieur (H2020 par exemple). La demande est donc le financement d'un ingénieur de recherche sur 2 ans, soit un total de l'ordre 96k€ (2k€ mensuels plus charges) ;
- par année, un stagiaire de M2 en statistique ou épidémiologie et un stagiaire de M2 en *Science et société* (gratification minimale de 3,6 € par heure), soit un total de 10k€ (stages de 6 mois) ;
- un logiciel communément employé dans le cadre des modèles causaux et des réseaux bayésiens est BayesiaLab (<http://www.bayesia.com/>). Seraient nécessaires une licence annuelle pour l'API permettant l'articulation avec les autres logiciels statistiques, essentiellement R, (*Bayesia Engine DEV-Academic SUSM 1-Year*) et deux licences annuelles du logiciel (*AcE SUSM 1-Year*), soit pour deux ans de l'ordre de 4k€ (chaque licence annuelle est de 490€ hors taxes).

6-Annexes

Bibliographie

- [1] Giroux É. Contribution à l'histoire de l'épidémiologie des facteurs de risque. *Revue d'Histoire des Sciences* 2011;**64**(2), 219-24.
- [2] Nwaru BI, Nurmatov U, Devereux G, Sheikh A. Potential confounders in the asthma–diet association: how causal approach could help? *Allergy* 2012(**67**):1461-3.
- [3] Fisher RA. *The design of experiments*. Oxford: Oliver and Boyd, 1935.
- [4] Collins H. *Tacit and explicit knowledge*. The University of Chicago Press Books, Chicago, 2012.
- [5] Parascandola M. The epidemiologic transition and changing concepts of causation and causal inference. *Revue d'Histoire des Sciences* 2011;**64**(2), 243-62.
- [6] Weed DL. On the use of causal criteria. *International Journal of Epidemiology* 1997;**26**, 1137-41.
- [7] Hill A. The environment and disease: association or causation. *Proceedings of the Royal Society of Medicine* 1965;**58**, 295-300.
- [8] Cartwright N. *Hunting causes and using them: approaches in philosophy and economics*. Cambridge: Cambridge University Press, 2007.
- [9] Ward AC. The role of causal criteria in causal inferences: Bradford Hill's "aspects of association". *Epidemiologic Perspectives and Innovations* 2009;**6**, 2.
- [10] Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000;**11**, 550-60.
- [11] Hume D. *An Enquiry Concerning Human Understanding*, 1748.
- [12] Suppes P. *A Probabilistic theory of causality*. Amsterdam: North-Holland Publishing Company, 1970.
- [13] Eells E. *Probabilistic causality*. Cambridge: Cambridge University Press, 1991.
- [14] Pearl J. An introduction to causal inference. *International Journal of Biostatistics* 2010;**6**, article 7.

- [15] Pearl J. *Causality: models, reasoning, and inference*. Cambridge: Cambridge University Press, 2000.
- [16] Lewis D. Causation. *Journal of Philosophy* 1973;**70**, 556-67.
- [17] Lewis D. Postscripts to "Causation". In *Lewis D. Philosophical Papers, Volume II*. Oxford: Oxford University Press, 1986, pp173-213.
- [18] Dawid AP. Counterfactuals, hypotheticals and potential responses: a philosophical examination of statistical causality. In *Causality and probability in the Sciences*, Eds Russo F, Williamson J. London: College Publications, 2007;**5**, 503-32.
- [19] Dawid AP. The role of scientific and statistical evidence in assessing causality. In *Perspectives on Causation*, Ed Goldberg R. Oxford: Hart Publishing, 2011, 133-47.
- [20] Spirtes P, Glymour C, Scheines R. *Causation, prediction, and search*. Cambridge: MIT Press (second edition), 2000.

CV du porteur de projet

Fonctions : PU-PH au Laboratoire de Biostatistique et Informatique Médicale et au Pôle de Santé Publique - Santé au Travail - Hygiène hospitalière (46.04 Biostatistiques, informatique médicale et technologies de communication).

Diplômes :

- 2015 - Nomination PU-PH
- 2013 - Master M2 Épistémologie, histoire des sciences et des techniques, Strasbourg
- 2013 - HDR *Des statistiques fréquentistes aux statistiques bayésiennes*, Strasbourg
- 2012 - Diplôme Inter-Universitaire de Pédagogie médicale, Strasbourg
- 2008 - Concours national des MCU-PH
- 2002 - Thèse *Évaluation de l'exposition professionnelle pour l'épidémiologie prospective*, Lyon
- 1998 - Concours National des Praticiens Hospitaliers
- 1996 - Doctorat en Médecine, Nancy (Lauréat de la Faculté), DES Santé Publique et Médecine Sociale

Activités de Recherche : Mon activité de recherche est organisée en trois grands axes qui s'entrecroisent en partie : modélisation géographique, méthodologie de la recherche clinique et de l'épidémiologie hospitalière et inférence bayésienne (notamment dans ses aspects historiques et épistémologiques). A ces axes, s'ajoutent des thèmes supplémentaires, notamment en cancérologie (Registre des Cancers du Haut-Rhin, Centre de gestion du dépistage du cancer colorectal en Alsace et le réseau français des registres du cancer, Francim). Je suis le porteur du projet *Se-REIN Échelle géographique et facteurs socio-économiques dans le traitement de suppléance de l'insuffisance rénale chronique*. Agence de biomédecine, 2014.

Publications dans des revues indexées : j'ai produit un total de 151 items (articles, communications, ...) dont 60 articles ou chapitres d'ouvrage. Parmi ces 60, 24 concernent les biostatistiques et le reste résulte de collaborations (recherche clinique, biologie, ...). L'index h est de 12.

Activités d'enseignement : Je suis responsable pédagogique du niveau 1 du C2i Santé (pour les études médicales et les orthophonistes) et du DU de Biostatistique *Des statistiques fréquentistes aux statistiques bayésiennes*. J'effectue de l'ordre de 90 heures d'équivalent-TD d'enseignement par année universitaire, dont les deux tiers dans le cadre des études des sciences médicales.

Encadrement d'étudiants : Le Département d'Information Médicale a l'agrément pour deux postes de DES Santé publique. J'ai assuré l'encadrement de plus de 20 étudiants : essentiellement pour le stage de M2 du Master Statistique et applications, Strasbourg (4), du Master Francilien de Santé Publique, Paris (2), du Master de Santé Publique - Environnement Grand Est (2). J'ai été directeur de 3 thèses d'exercice de médecine et actuellement directeur de trois thèses de troisième cycle (dont deux en co-direction).

Activités hospitalières : En tant qu'un des méthodologistes du Groupe Méthode en Recherche Clinique des Hôpitaux Universitaires de Strasbourg, j'assure une centaine de consultations de statistique par an, dont 80% est de l'épidémiologie hospitalière et 20% de la recherche clinique. Je suis le méthodologiste d'un PRME (2014), un PHRIP (2014), deux PHRC nationaux (2010 et 2011), deux PHRC inter-régionaux (2013) et trois appels à projet Jeunes chercheurs (2012 et 2014). Dans les autres travaux hospitaliers, on peut citer les études d'évaluation de la mise en place du système d'information médicalisé, les études sur les données issues du PMSI, la fouille de données dans les données de santé.

Autres activités : je suis membre du Comité de protection des personnes Est-4. Je suis relecteur régulier pour *Bayesian Analysis*, *PLOS One*, *Journal of Applied Statistics*, *International Journal of Environmental Research and Public Health*, membre du Comité Éditorial du *Journal d'Économie et de Gestion Médicales* et membre du bureau du groupe Enseignement à la Société Française de Statistique.

Réponse au rapport du conseil scientifique

Le rapport du conseil scientifique du LabEx IRMIA (transmis le 30 janvier 2016) est le suivant :

L'objectif général du projet consiste à réfléchir aux fondements philosophiques et épidémiologiques/biostatistiques de la causalité et de la modélisation causale, avec un ancrage tout particulier sur l'historique de ces approches et leur historiographie. L'approche reproche essentiellement sur la constitution d'un réseau scientifique international disciplinaire impliquant des collaborateurs de Strasbourg, Cambridge et Cagliari.

Il s'agit selon moi d'un dossier intéressant, pour plusieurs raisons :

1- L'étude des phénomènes causaux est un domaine en pleine expansion, dont l'application essentielle se situe dans les études biostatistiques dites observationnelles. La question typique est celle de l'efficacité d'un traitement médical sur des sujets n'ayant pas pu subir de randomisation au préalable. On cherche alors à mesurer l'efficacité du traitement indépendamment de toute facteur ayant pu influencer l'appartenance à un groupe ou à l'autre.

2- Le projet est véritablement pluridisciplinaire et international.

3- Au delà de la biostatistique, de la médecine et de la sociologie, la causalité est le notions connexes seront incontestablement amenées à jouer un rôle essentiel dans les études sur les données à grande échelle, complexes et en ultra-grande dimension (registre big data).

4- Les collaborateurs sont expérimentés et possèdent des sensibilités scientifiques très variées.

D'un point de vue plus négatif, on peut peut être s'interroger sur d'une part l'adéquation entre le projet et les axes du labex IRMIA et, d'autre part, sur l'absence de partenariat clair avec des collègues mathématiciens de l'IRMA.

En conclusion, mon avis est favorable à condition que le projet s'inscrive de façon claire dans les priorités de l'IRMIA et/ou qu'il implique d'une manière plus explicite des statisticiens de l'IRMA.

➔ Le projet INCEPTION a été modifié pour répondre aux deux points négatifs, justifiés, du rapport :

- la section 4 du projet a été re-écrite et pointe maintenant plus clairement certains axes scientifiques du LabEx (<http://labex-irmia.u-strasbg.fr/spip.php?article6&lang=fr>) pour montrer comment INCEPTION s'y insère ;
- l'implication explicite des statisticiens de l'IRMA apparaît maintenant clairement par la modification de la section 5 et par l'ajout d'une fiche de poste pour l'ingénieur de recherche. Désormais il est clair que son encadrement sera conjoint entre le porteur du projet et un mathématicien de l'IRMA (Mr Frédéric Bertrand).

Fiche de poste

Désignation du poste : INGENIEUR DE RECHERCHE

Mission / Activités

L'ingénieur de recherche sera en charge du traitement statistique de jeux de données réelles à l'aide de modèles causaux. Il aura également un rôle de coordination des interactions et de communication avec les praticiens et les autres membres du projet INCEPTION. Il rédigera des notes méthodologiques.

Ses activités principales seront :

- Choisir et mettre en œuvre des méthodes d'exploration de données et d'analyse de la causalité
- Évaluer et valider l'application de méthodes à un problème statistique
- Former et assurer le transfert des connaissances et des savoir-faire aux praticiens

Des activités associées lui incomberont aussi :

- Diffuser et valoriser les méthodes et outils statistiques développés
- Participer à des activités de recherche
- Suivre et intégrer l'évolution des concepts dans les domaines des méthodes d'analyse et modélisation statistique

Compétences

1. Savoirs généraux théoriques ou disciplinaires
 - a. Connaissance approfondie des principales méthodes d'analyse ou de modélisation statistique
 - b. Une connaissance en inférence bayésienne serait un avantage
2. Savoir-faire opérationnels
 - a. Connaissance approfondie du logiciel R
3. Compétences linguistiques
 - a. Anglais : Expression écrite et orale : bon niveau
 - b. Compréhension écrite et orale : bon niveau

Environnement et formations

- L'activité sera localisée au Laboratoire de Biostatistique et Informatique Médicale de la Faculté de Médecine de l'Université de Strasbourg, au sein du Laboratoire ICube (UMR 7357).
- Diplôme exigé : un master est exigé, un doctorat est souhaité.
- Formation et expérience professionnelle souhaitables : une expérience professionnelle de plusieurs années dans le domaine de la modélisation statistique est souhaitable.

Déroulement du travail (24 mois)

1. 1er temps (3-4 mois)
 - a. Travail de bibliographie et appropriation du code existant
 - b. En parallèle ou au préalable choix de jeux de données typiques.
2. 2ème temps (12-15 mois) : application à cinq jeux de données de nature variée en complexité croissante (augmentation du nombre de variables explicatives, introduction d'un recueil répété dans le temps, passage au big data). Outre l'ajustement du modèle, plusieurs autres points feront l'objet d'une étude :
 - a. Évaluation et limitations des différents modèles causaux.
 - b. Adéquation, choix et validation de modèle.
 - c. Utilisation et influence des données a priori dans la construction du modèle.
3. 3ème temps (3mois) : rédaction d'une synthèse, participation à la préparation des appels d'offre et d'un colloque.

S'ajoute à ces travaux la coordination des interactions entre chercheurs et la communication autour du projet (2 mois diffus).