

RAPPORT D'ÉVALUATION DE L'UNITÉ

BioSTM - Biostatistique, traitement et
modélisation des données biologiques

SOUS TUTELLE DES ÉTABLISSEMENTS ET ORGANISMES :

Université Paris Cité - UP Cité

CAMPAGNE D'ÉVALUATION 2023-2024
VAGUE D



Au nom du comité d'experts :

Sophie Donnet, Présidente du comité

Pour le Hcéres :

Stéphane Le Bouler, président par intérim

En application des articles R. 114-15 et R. 114-10 du code de la recherche, les rapports d'évaluation établis par les comités d'experts sont signés par les présidents de ces comités et contresignés par le président du Hcéres.

Pour faciliter la lecture du document, les noms employés dans ce rapport pour désigner des fonctions, des métiers ou des responsabilités (expert, chercheur, enseignant-chercheur, professeur, maître de conférences, ingénieur, technicien, directeur, doctorant, etc.) le sont au sens générique et ont une valeur neutre.

Ce rapport est le résultat de l'évaluation du comité d'experts dont la composition est précisée ci-dessous. Les appréciations qu'il contient sont l'expression de la délibération indépendante et collégiale de ce comité. Les données chiffrées de ce rapport sont les données certifiées exactes extraites des fichiers déposés par la tutelle au nom de l'unité.

MEMBRES DU COMITÉ D'EXPERTS

Présidente : Mme Sophie Donnet, Inrae Paris

Expert : M. Bertrand Maury, Université Paris-Saclay (représentant du CNU)

REPRÉSENTANT DU HCÉRES

M. Philippe Elbaz-Vincent

REPRÉSENTANTS DES ÉTABLISSEMENTS ET ORGANISMES TUTELLES DE L'UNITÉ DE RECHERCHE

M. Jean-Louis Beaudoux, Université Paris Cité
M. Xavier Declèves, Université Paris Cité
Mme Christine Guillard, Université Paris Cité
M. Michel Vidal, Université Paris Cité

CARACTÉRISATION DE L'UNITÉ

- Nom : Biostatistique, traitement et modélisation des données biologiques
- Acronyme : BIOSTM
- Label et numéro : URP 7537
- Composition de l'équipe de direction : ROZENHOLC Yves

PANELS SCIENTIFIQUES DE L'UNITÉ

ST Sciences et technologies
ST1 Mathématiques

THÉMATIQUES DE L'UNITÉ

L'unité BioSTM est une unité mono-équipe qui rassemble des chercheurs en modélisation mathématique et apprentissage statistique pour les données des sciences pharmaceutiques et plus généralement de la santé. En effet, la recherche fondamentale ou le développement de traitements thérapeutiques génèrent des données hétérogènes dont la nature et le volume évoluent très rapidement au rythme des nouvelles techniques biologiques.

L'unité a donc une double mission de 1/ expertise en apprentissage statistique et modélisation mathématique et de 2/ développement méthodologique pour des questions issues des sciences pharmaceutiques. Ses travaux s'appuient sur une forte expertise non seulement en apprentissage statistique, traitement du signal et systèmes dynamiques, mais aussi en biologie, chimie ou pharmacie que possèdent certains membres.

HISTORIQUE ET LOCALISATION GÉOGRAPHIQUE DE L'UNITÉ

L'unité a été créée le 01/01/2019 afin de rassembler l'ensemble des chercheurs et enseignants-chercheurs de la faculté de Pharmacie, spécialistes en apprentissage statistique et modélisation pour des données des sciences pharmaceutiques, et jusque-là éparpillés dans différents laboratoires de biologie et de chimie. Ce regroupement visait à favoriser une meilleure synergie entre chercheurs en mathématiques appliquées autour d'un objectif commun d'analyse de données appliquées dans des contextes différents. En effet, la nature des données et les méthodes d'apprentissage évoluant très rapidement, cette organisation centrée sur l'outil méthodologique cherche à garantir une meilleure dynamique d'innovation dans l'analyse de données des sciences pharmaceutiques.

L'unité est implantée dans les locaux de la Faculté de Pharmacie de l'université Paris Cité située au 4 av. de l'Observatoire à Paris.

ENVIRONNEMENT DE RECHERCHE DE L'UNITÉ

Institutionnellement, la faculté de pharmacie de Paris à laquelle est rattachée l'unité est une composante de la faculté de Santé de l'Université Paris Cité.

L'unité s'inscrit dans le riche paysage de la recherche francilienne en modélisation mathématique et apprentissage statistique. Elle se distingue des autres laboratoires du champ disciplinaire par son domaine d'application, ici les sciences pharmaceutiques, par les compétences multidisciplinaires internes qu'elle revendique et par sa volonté de s'inscrire dans le caractère appliqué et applicable de ses travaux.

EFFECTIFS DE L'UNITÉ : en personnes physiques au 31/12/2022

Catégories de personnel	Effectifs
Professeurs et assimilés	2
Maîtres de conférences et assimilés	6
Directeurs de recherche et assimilés	0
Chargés de recherche et assimilés	0
Personnels d'appui à la recherche	0
Sous-total personnels permanents en activité	8

Enseignants-chercheurs et chercheurs non permanents et assimilés	0
Personnels d'appui non permanents	0
Post-doctorants	0
Doctorants	1
Sous-total personnels non permanents en activité	1
Total personnels	9

RÉPARTITION DES PERMANENTS DE L'UNITÉ PAR EMPLOYEUR : en personnes physiques au 31/12/2022. Les employeurs non tutelles sont regroupés sous l'intitulé « autres ».

Nom de l'employeur	EC	C	PAR
UNIVERSITÉ PARIS-CITÉ	8	0	0
Total personnels	8	0	0

AVIS GLOBAL

Au cours de la période évaluée, l'unité a rempli l'objectif qu'elle s'était fixé, celui de rassembler les enseignants-chercheurs (EC) en sciences des données de la faculté de pharmacie dans une unique entité afin de rompre l'isolement scientifique de ses chercheurs et de créer une dynamique de recherche collaborative.

L'unité est maintenant composée de neuf (fin 2023) enseignants-chercheurs en mathématiques appliquées et apprentissage statistiques. Certains membres ont de plus une double compétence en pharmacie, chimie ou biologie. Les recrutements récents ont ajouté des compétences nouvelles à l'unité (analyse de données omiques pour maladies complexes ou informatiques et transport optimal).

L'unité fonde son activité sur des sollicitations extérieures, issues d'unités d'autres domaines qui requièrent une expertise sur l'analyse de leurs données biologiques, telles que données génomiques, images, exposomes, suivi de cohortes, essais cliniques, spectres. En fonction de la question posée, la réponse consiste en une activité de service ou bien d'innovation méthodologique, comme mis en avant dans le portfolio. Toute collaboration aboutit à une publication, généralement dans un journal du domaine d'application (une dizaine par an pour l'unité). Par choix stratégique, l'unité ne met pas en avant son leadership sur une méthodologie statistique ou mathématique précise, mais bien sur le domaine d'application que sont les données de santé. Ainsi, l'unité se positionne sur la scène internationale en apportant son expertise mathématique pour de grandes avancées en biologie, notamment autour de l'étude du phénomène de la pléiotropie (présentée dans le portfolio).

La grande activité et le succès des membres dans le dépôt de projets (dix projets depuis 2019, financés en moyenne à hauteur de 58 k€) permettent à l'unité d'avoir les ressources nécessaires à son activité de recherche et d'encadrement. Le dépôt prévu d'un projet européen est révélateur du dynamisme collectif et de l'ambition portée par l'unité.

Le choix de l'unité de publier essentiellement dans des journaux de biologie et de médecine notamment, limite sa capacité à faire connaître ses travaux auprès de la communauté «stat/math», bien que cet état de fait soit tempéré par la participation de l'unité à des conférences nationales et internationales de cette communauté et que BioSTM est largement impliquée dans le GDR «Statistique et Santé». Une telle stratégie prive néanmoins la communauté «stats/math», voire la communauté scientifique en général, des innovations développées par BioSTM pour les données biologiques et potentiellement transférables à d'autres domaines que la santé humaine. En outre, la construction en cours du site internet institutionnel permettra de faire connaître la composition de l'unité, ses thématiques de recherche ou de retrouver rapidement ses packages et publications. La trajectoire de l'unité est en droite ligne avec sa stratégie et son positionnement actuel. Elle souhaite poursuivre son activité. Il semble important au comité que l'unité poursuive son activité scientifique à ce niveau et augmente sa visibilité scientifique aux échelles régionales et nationales notamment auprès de la communauté statistique et mathématique. Cette visibilité semble importante dans le contexte futur d'intégration de l'unité comme équipe d'une unité de service de la faculté.

ÉVALUATION DÉTAILLÉE DE L'UNITÉ

A - PRISE EN COMPTE DES RECOMMANDATIONS DU PRÉCÉDENT RAPPORT

Non applicable ici.

B - DOMAINES D'ÉVALUATION

DOMAINE 1 : PROFIL, RESSOURCES ET ORGANISATION DE L'UNITÉ

Appréciation sur les objectifs scientifiques de l'unité

Au cours de la période évaluée, l'unité s'était fixée comme dessein de rassembler les EC en sciences des données de la faculté de pharmacie dans une unique entité afin de créer une dynamique et de rompre l'isolement scientifique de ses chercheurs. L'objectif de la période passée était donc de créer une dynamique au sein de ce groupe. L'activité scientifique décrite dans le document d'autoévaluation, les entretiens avec le personnel et avec les tutelles montrent que ces objectifs ont été atteints. Précisons que la part essentielle de l'activité de l'unité consiste à répondre à des demandes de l'extérieur, être capable de répondre à ces demandes de façon satisfaisante constitue l'objectif essentiel, et assumé, de l'unité.

Appréciation sur les ressources de l'unité

L'ensemble de ses ressources financières lui permet de financer des stages, contrats doctoraux et post-doctoraux, d'équiper ses membres en matériel informatique et d'assister aux conférences nécessaires à leur activité scientifique. Les locaux sont de taille en adéquation avec l'effectif actuel de l'unité. Ils nécessitent quelques travaux de rénovation qui semblent en cours. En termes de ressources humaines sur postes permanents, l'unité a bénéficié de deux recrutements de maîtres de conférences (MCF) depuis sa création. La temporalité des recrutements est conditionnée par les besoins en enseignement. L'unité dispose de ressources limitées en termes de support administratif (une gestionnaire partagée avec une autre unité), et inexistantes en termes de personnel support à la recherche (pas d'ingénieur système ni d'ingénieur de recherche), malgré le caractère essentiel du calcul intensif dans ses activités.

Appréciation sur le fonctionnement de l'unité

L'unité étant de très petite taille, elle n'a pas mis en place d'outils formels de fonctionnement. L'ensemble des personnels est localisé dans une unité d'espace, ce qui facilite la communication administrative et scientifique. Le groupe semble soudé, les membres font état d'une discussion aisée et les décisions semblent collectives, sans cloisonnement entre les différentes catégories de personnel. Les doctorants sont tout à fait intégrés au groupe et se sentent écoutés et soutenus.

1 / L'unité s'est assigné des objectifs scientifiques pertinents.

Points forts et possibilités liées au contexte

La science de données évoluant très rapidement, le modèle ancien de statisticiens dans des unités de recherche en biologie ou sciences du vivant n'est plus pertinent puisqu'il isole le modélisateur de sa communauté et le prive de montée en compétence. Par conséquent, l'idée de regrouper les mathématiciens appliqués de la faculté de pharmacie dans une unité méthodologique spécialisée était tout à fait pertinente. Les interactions entre ces chercheurs (visibles au travers de publications et des projets déposés) prouvent que cette création est un succès. Les entretiens avec les EC ont permis de constater la satisfaction de ces personnes

de ce point de vue. Les EC présents à l'origine soulignent les bénéfices de cette création et font part d'une grande satisfaction scientifique, intellectuelle et humaine. Les nouveaux recrutés semblent apprécier la variété de compétences présente dans ce groupe. Par ailleurs, à l'échelle de la faculté, l'unité semble reconnue pour ses compétences transverses en sciences de données. Cet aspect a été souligné par les représentants des tutelles et est étayé par le grand nombre de travaux pluridisciplinaires publiés (voir analyse de la production scientifique).

Par son positionnement, le fonctionnement de l'unité repose essentiellement sur l'existence de collaborations avec d'autres unités de la faculté. Ces collaborations sont installées depuis très longtemps et l'organisation entre les membres est telle que les collaborations individuelles bénéficient aux autres membres de l'unité, en particulier aux nouveaux recrutés.

En outre, grâce à sa position thématique unique au sein de l'université, on aurait pu craindre une sur-sollicitation des membres de l'unité. Il semble que là encore, le groupe soit organisé de telle façon que ses membres ne souffrent pas d'une telle pression. De plus, ils semblent en mesure de sélectionner les collaborations de recherche en fonction des compétences ou du challenge ou intérêt méthodologique que les sujets représentent.

Points faibles et risques liés au contexte

L'unité tient à s'afficher comme généraliste d'un point de vue méthodologique, mais spécialiste dans les domaines d'application en se concentrant sur des données de santé (pharmacie, médecine, épidémiologie). À ce titre, l'unité ne met en avant aucune spécialité méthodologique précise, que ce soit dans le document d'autoévaluation ou dans l'introduction au portfolio. Par ailleurs, les travaux sont essentiellement publiés dans des journaux des domaines d'applications (voir domaine D4 de ce rapport). Cette stratégie généraliste d'une part et de diffusion des travaux à destination de publics non mathématiciens présente le risque de déconnecter l'unité de cette communauté. Ce thème sera évoqué dans le domaine D3 (Production scientifique).

2/ L'unité dispose de ressources adaptées à son profil d'activités et à son environnement de recherche et les mobilise.

Points forts et possibilités liées au contexte

L'unité dispose de ressources par dotation de l'université (74K€ pour la période écoulée) et de ressources propres obtenues sur appels à projets (ANR par exemple). Durant la période, l'unité a obtenu 595 k€ à répartir sur les années des projets.

Grâce à ces financements (dotation université et contrats sur projet, ces derniers constituant l'essentiel des ressources), l'unité parvient à s'équiper matériellement à la hauteur des besoins. Concernant les équipements numériques, elle dispose de ses propres serveurs, ce qui lui permet de ne pas faire appel à des ressources de calcul extérieures : un serveur de calcul haute performance (128 cœurs), un GPU de 12 000 cœurs et un NAS d'une capacité de 40 To de stockage. Les ordinateurs sont régulièrement changés. Les membres de l'unité disposent d'écrans à domicile financés par l'unité et de matériel de bureau ergonomique et confortable. Le positionnement de l'unité en sciences des données dans un but pluridisciplinaire lui permet d'accéder à des financements de multiples guichets.

Les locaux sont de taille raisonnable étant donnés les effectifs actuels de l'unité. Les bureaux sont de taille standard pour la région parisienne intra-muros : certains membres partagent leurs bureaux, d'autres ont un bureau personnel. Les locaux sont bien éclairés bien que légèrement sous le niveau du sol, et avec une vue agréable. L'unité s'est équipée d'un coin cuisine qui permet à ses membres de déjeuner ensemble. Ce coin cuisine sert aussi de salle de réunion. Les locaux sont à la même adresse que les lieux d'enseignement. Certaines parties nécessitent des réparations et rénovations, qui devraient intervenir au printemps 2024.

D'un point de vue ressources humaines, l'unité a obtenu récemment un poste de MCF, augmentant ainsi son effectif de 12,5%. L'obtention de postes est directement reliée aux besoins en enseignements. La tutelle s'attache à rendre stable la masse salariale, en remplaçant les départs sans réduction de personnel.

Points faibles et risques liés au contexte

L'unité est de très petite taille (neuf membres permanents en janvier 2024) et par conséquent elle rencontre les difficultés et risques liés à sa petite taille, notamment la difficulté de bénéficier d'une cellule d'appui à la recherche. L'activité de service en sciences de données portée par l'unité pour l'instant à la charge des EC pourrait mériter le recrutement d'un ingénieur (IR ou IE). Concernant la gestion, l'unité ne dispose pas de

secrétariat et sa gestionnaire est partagée avec une autre unité et semble en souffrir. Enfin, les ressources de calculs sont gérées par le directeur d'unité (DU) lui-même, ce qui pourrait ou devrait être fait par un informaticien de métier.

Le DU souligne une diminution de 20 % de la dotation de l'université, qui correspond à une mesure globale affectant l'ensemble des unités. Cela ne semble pas mettre en difficulté les EC, notamment car l'unité cherche activement des ressources en déposant des projets du type ANR ou même ERC. Il est important de noter que bien que l'unité soit en mesure de trouver des financements auprès de multiples guichets (grâce au caractère pluridisciplinaire de sa recherche), cette activité de recherche de financement est chronophage et vient empiéter le temps consacré à la recherche en tant que telle.

3/ Les pratiques de l'unité sont conformes aux règles et aux directives définies par ses tutelles en matière de gestion des ressources humaines, de sécurité, d'environnement, de protocoles éthiques et de protection des données ainsi que du patrimoine scientifique.

Points forts et possibilités liées au contexte

Grâce à son cœur de métier et sa taille, l'unité n'est pas très concernée par ces points. L'unité ne produit pas de données biologiques, aucune situation de travail ne génère de risque matériel particulier. Les seuls risques pourraient être psychosociaux. Les entretiens avec les doctorants et post-doctorants ont montré qu'ils étaient informés des ressources à leur disposition en cas de situation conflictuelle. Par ailleurs, ils ont identifié les permanents de l'unité comme première personne de conseil, ce qui témoigne du climat de confiance régnant au sein du groupe. Malgré l'absence de réunions ou de commissions programmées à intervalles réguliers, l'information circule de façon fluide au sein de l'unité.

Points faibles et risques liés au contexte

Sans objet.

DOMAINE 2 : ATTRACTIVITÉ

Appréciation sur l'attractivité de l'unité

Malgré sa taille réduite, l'unité bénéficie d'une bonne attractivité aux niveaux local et national, du fait de son positionnement pluridisciplinaire original et de sa production scientifique, qui met la méthodologie au service de l'application. L'unité s'inscrit dans des thématiques médicales de pointe.

- 1/ L'unité est attractive par son rayonnement scientifique et s'insère dans l'espace européen de la recherche.*
- 2/ L'unité est attractive par la qualité de sa politique d'accompagnement des personnels.*
- 3/ L'unité est attractive par la reconnaissance de ses succès à des appels à projets compétitifs.*
- 4/ L'unité est attractive par la qualité de ses équipements et de ses compétences techniques.*

Points forts et possibilités liées au contexte pour les quatre références ci-dessus

L'unité est attractive par l'originalité de son positionnement dans le paysage français. Ce positionnement pluridisciplinaire est attractif pour des profils avec doubles compétences (apprentissage et pharmacie ou biologie), dont l'originalité est assez peu reconnue et valorisée dans le système français classique.

Les jeunes EC bénéficient d'un service d'enseignement réduit l'année de leur recrutement. Aucune remarque n'a été faite sur une éventuelle surcharge des heures d'enseignements.

D'un point de vue recherche, les EC expérimentés qui sont à l'origine des collaborations impliquent systématiquement les jeunes EC dans les nouvelles collaborations : ils font ainsi profiter les EC débutants de leur réseau personnel.

Plusieurs projets sont en cours dont notamment une ANR JCJC portée par une jeune recrutée, épaulée par un EC plus expérimenté. La tutelle assure avoir les moyens humains pour aider et accompagner ses EC à déposer des projets européens. Ces succès aux appels à projets permettent de financer des thèses, missions et matériel.

Les locaux, idéalement situés au cœur de Paris, sont dans un état correct. Tous les membres de l'unité y sont regroupés. Les bureaux sont à la même adresse que les activités d'enseignement. L'unité s'est dotée d'une puissance de calcul qui lui permet d'être autonome. Les conditions matérielles (bureau, machine) sont bonnes.

L'attractivité de l'unité sur tous ces critères est également attestée par le flux important de doctorants et de stagiaires de master 2, qui bénéficient d'un bon environnement de travail et d'un encadrement suivi.

Points faibles et risques liés au contexte pour les quatre références ci-dessus

Le positionnement pluridisciplinaire peut aussi être un risque, notamment car il peut être difficile pour l'unité de clarifier son positionnement dans le paysage français. Cet inconvénient doit être source de vigilance notamment pour les personnels en début de carrière (doctorants, post-doctorants et jeunes EC).

Par ailleurs, la petite taille de l'unité peut paraître comme un point faible en termes d'animation scientifique. Cette taille ne permet pas d'animer un séminaire par exemple. Enfin, la taille rend l'unité fragile face à des départs ou absences éventuels. Tout départ non anticipé entraînerait une perte de compétence.

DOMAINE 3 : PRODUCTION SCIENTIFIQUE

Appréciation sur la production scientifique de l'unité

L'unité a une production scientifique de qualité (environ une dizaine d'articles par an) portée par l'ensemble de ses membres. À cette production d'articles s'ajoutent la création et la maintenance de packages. Certains articles écrits par des membres de l'unité ont été cités plusieurs milliers de fois. Cependant, la majorité des travaux sont publiés dans des journaux hors statistiques ou mathématiques, entraînant un manque de visibilité pour ces innovations méthodologiques auprès de la communauté «stats/math». Cet aspect est dommageable pour la communauté «stats/math» et pour la communauté scientifique en général puisqu'on peut penser que les méthodes développées dans le cadre de l'analyse de données issues des sciences de la santé humaine pourraient être appliquées à d'autres domaines (santé animale, santé des plantes, environnement notamment).

Le portfolio présenté est composé de cinq articles et d'une conférence invitée. Il aurait été intéressant pour chaque article de le présenter comme soit une activité de service (application d'une méthode existante et support d'expertise) ou soit de recherche. Dans ce dernier cas, une description de l'innovation méthodologique d'un point de vue statistique aurait permis à la lecture du dossier de mieux situer les travaux de l'unité dans le paysage de l'apprentissage statistique. Cette clarification a été faite au cours des entretiens avec le comité et a permis d'appréhender la qualité et la pertinence des avancées en termes d'analyse de données.

1/ La production scientifique de l'unité satisfait à des critères de qualité.

2/ La production scientifique de l'unité est proportionnée à son potentiel de recherche et correctement répartie entre ses personnels.

3/ La production scientifique de l'unité respecte les principes de l'intégrité scientifique, de l'éthique et de la science ouverte. Elle est conforme aux directives applicables dans ce domaine.

Points forts et possibilités liées au contexte pour les trois références ci-dessus

Le portfolio et l'ensemble des publications de l'unité ont permis de mesurer le large spectre des thématiques abordées dans l'unité. Ces thématiques incluent l'apprentissage statistique ou automatique, notamment pour des tâches de classification et de prédiction, l'inférence bayésienne ou encore l'inférence causale. Le traitement du signal est aussi une thématique abordée par l'unité.

Le portfolio présenté par l'unité incluait cinq articles publiés dans des revues internationales à comité de lecture. Elle visait ainsi à mettre en avant la diversité et l'originalité des réponses techniques apportées à des questions médicales.

Les deux exposés scientifiques lors des entretiens ont mis en avant des innovations méthodologiques pour l'analyse d'expressions de gènes. Ainsi, un premier exposé portait sur la pléiotropie c'est-à-dire l'effet d'un variant génétique sur plusieurs caractères phénotypiques. Ces travaux sur la pléiotropie viennent compléter les nombreuses études d'association à l'échelle du génome qui ont permis d'identifier des dizaines de milliers de variantes génétiques associées à des caractéristiques complexes (prises une par une). Dans les travaux présentés, une inférence causale des dits effets est proposée, mettant en concurrence plusieurs scénarios biologiques traduits en hypothèses mathématiques. L'obtention de résultats requiert l'utilisation d'outils d'apprentissage automatique, mais repose aussi sur une double compétence en biologie (au sens large) et mathématiques appliquées (modélisation et apprentissage) afin d'avoir une compréhension fine des conséquences des résultats.

Le deuxième exposé portait sur le traitement des données compositionnelles : en effet, les données d'expression de gènes sont obtenues à quantité totale de RNA fixée, rendant les expressions de chaque gène dépendantes. Afin de prendre en compte cette relation entre les variables, la pratique classique consiste à renormaliser les quantités par rapport à un ou plusieurs gènes de référence, dont la quantité varie peu au sein de la population étudiée et qui ne sont pas impliquées dans le mécanisme de régulation à l'étude. Le choix de cette quantité de référence est critique pour l'interprétabilité des résultats biologiques et notamment conditionne la décision à propos du rôle éventuel d'un gène dans une pathologie. Les travaux présentés ont mis en avant une méthode innovante permettant un choix éclairé de ces gènes de référence. La méthode consiste à construire un réseau entre gènes par des tests statistiques d'équivalence et identifier l'ensemble des gènes de référence à partir de la topologie du réseau.

Dans les deux cas, les motivations et les outils produits (notamment graphiques) soulignent une fois de plus que l'unité poursuit ses objectifs en fournissant des réponses mathématiquement fondées et applicables à des questions biologiques d'actualité.

L'extraction de la base HAL permet de comptabiliser un total de 25 publications internationales dans des revues à comité de lecture durant la période 2020-2022 (l'unité n'est référencée sur HAL qu'à partir de 2020). En sus, la liste fournie par l'unité permet de comptabiliser treize publications pour l'unité en 2019 par les membres présents dans l'unité.

Le nombre de publications est stable d'une année sur l'autre. Parmi ces publications, environ 25 % impliquent au moins deux membres de l'unité. L'intégralité de ces publications sont cosignées avec des scientifiques d'autres domaines. Pour environ un tiers des publications, les membres de l'unité sont premiers, deuxièmes ou derniers auteurs (et identifiés comme tels dans la liste fournie par l'unité). L'ensemble des membres de l'unité participe à cette production. Le nombre de publications attribuées aux doctorants (2) est évidemment réduite en raison de la taille et de la jeunesse de l'unité, mais sur ces deux publications les doctorants sont premiers auteurs.

Le ratio articles de journaux / actes de conférences montre que l'unité privilégie les publications dans des journaux. Certains travaux sont présentés à des conférences nationales ou internationales. La plupart des journaux de publications appartiennent aux domaines de la biologie, médecine ou pharmacologie. Plusieurs articles impliquent des chercheurs appartenant à des laboratoires à l'étranger. L'unité n'affiche pas de politique particulière de choix de revues, cependant au moins la moitié des publications sont en Open Access lisible sur Hal.

Certaines publications bénéficient d'une très bonne visibilité, notamment en termes de citations. À titre d'exemple, un article récent portant sur la pléiotropie a déjà été cité par plusieurs milliers de chercheurs.

En sus de ces publications scientifiques, l'unité produit et maintient des packages R ou Python. En 2022, l'unité maintenait neuf packages dont cinq sont des packages R référencés et disponibles sur le CRAN, deux sont des packages R disponibles sur GitHub et un package Python. La production de packages liés aux publications permet à l'unité d'assurer la diffusion des méthodes proposées, mais aussi la reproductibilité des résultats obtenus lorsque les données peuvent être rendues publiques.

Il est à noter que cette dynamique de production scientifique pure s'accompagne d'une grande implication de l'unité dans des organes d'évaluation (dont comités de sélection, jurys de thèse) et d'animation de la recherche (par exemple au sein du GDR «Statistique et Santé»).

Points faibles et risques liés au contexte pour les trois références ci-dessus

Bien que l'unité participe à des conférences en apprentissage statistique (par ex. SfdS, ISBA), les publications sont majoritairement dans des revues de biologie et médecine. Dans ce dernier cas, les innovations et réponses apportées en termes d'apprentissage statistique ne sont en général pas ou peu mises en avant dans le titre de l'article. Le choix de publier majoritairement dans des revues hors communauté «stats/math» appliquées réduit la visibilité des travaux de l'unité auprès de cette communauté. Ce manque de visibilité sur les méthodes d'apprentissage innovantes est dommageable pour la communauté «stats/math» dans la mesure où les données traitées sont les mêmes qu'en santé animale, santé des plantes ou médecine. Par conséquent, ces domaines ne peuvent pas profiter des innovations proposées par l'unité.

En outre, l'absence de publication des avancées méthodologiques dans des revues méthodologiques ne permet pas de valoriser et de mettre en évidence l'activité de recherche qui est masquée par les activités de service. Ce manque de visibilité auprès de la communauté «stats/math» au travers des publications se double d'un manque de visibilité institutionnelle à l'extérieur de la Faculté de pharmacie de UPC, bien que ces points ne soient pas vus comme des faiblesses par l'unité. En effet, lors des entretiens, l'unité ne possédait pas de portail HAL ni de site web permettant d'identifier son positionnement et sa composition en termes de personnels (la page web est depuis en cours de construction).

Concernant les interactions au sein de l'unité, bien que le groupe souligne des échanges scientifiques constants depuis la création, cette politique ne se traduit pas encore par des signatures de co-publications (ce qui paraît normal). Cependant, dans la mesure où les collaborations les plus fructueuses avec les autres unités de la faculté reposent essentiellement sur les membres les plus expérimentés, on peut craindre un risque, à long terme, de voir ces collaborations disparaître au départ de ces chercheurs. Ce risque pourrait être limité en augmentant les collaborations intra-unité.

DOMAINE 4 : INSCRIPTION DES ACTIVITÉS DE RECHERCHE DANS LA SOCIÉTÉ

Appréciation sur l'inscription des activités de recherche de l'unité dans la société

L'essentiel des activités scientifiques de l'unité a une visée applicative clairement identifiée, répondant en particulier à des besoins réels en termes de santé publique. Les activités explicitement tournées vers le grand public et la société se cristallisent autour du projet Covid-19 Museum, lancé par des membres de l'unité, qui a vocation à recueillir des «traces» de diverses natures de la pandémie récente.

- 1/ L'unité se distingue par la qualité et la quantité de ses interactions avec le monde non-académique.*
- 2/ L'unité développe des produits à destination du monde culturel, économique et social.*
- 3/ L'unité partage ses connaissances avec le grand public et intervient dans des débats de société.*

Points forts et possibilités liées au contexte pour les trois références ci-dessus

L'unité travaille sur des problématiques de recherche finalisée dont l'objet est la santé humaine. Par conséquent, par essence même, les travaux de l'unité ont un impact sur la société. En termes de médiation ou d'aide aux politiques publiques, l'activité de l'unité est concentrée sur le Covid-19 Museum qui vise à collecter les traces numériques de la pandémie et à partager des outils d'analyse de ces données. Dans ce cadre, l'unité qui porte le projet a donné un grand nombre de conférences et a acquis les ressources matérielles nécessaires pour héberger les données.

Points faibles et risques liés au contexte pour les trois références ci-dessus

La taille réduite de l'équipe rend difficile une participation massive à des activités de communications auprès du grand public, qui ne pourrait se faire qu'au dépend de la disponibilité pour les activités de recherche. Il pourrait néanmoins être pertinent de présenter une vitrine en ligne (pages web personnelles ou collectives) pour mettre en valeur, de façon accessible au plus grand nombre, les contributions les plus impactantes.

La mise en place du projet Covid-19 Museum témoigne d'une volonté affirmée de toucher un public au-delà des chercheurs et experts, et la place que ce projet occupe dans le rapport d'autoévaluation témoigne de l'importance accordée par certains membres de l'unité à ce projet. Néanmoins, les liens entre ce projet et l'unité ne sont pas affichés de façon très claire, et il est difficile de mesurer actuellement l'impact effectif de ce projet sur le grand public.

ANALYSE DE LA TRAJECTOIRE DE L'UNITÉ

D'un point de vue scientifique, l'unité, au travers de son directeur, affirme un positionnement futur plus qu'une trajectoire scientifique. Le groupe propose de continuer à s'afficher généraliste en développement des méthodes de sciences de données pour les données issues de la pharmacie et ne souhaite pas mettre en avant un sous-ensemble de compétences méthodologiques. L'unité se place ainsi à l'affut de toute opportunité de collaboration interdisciplinaire, certaines relevant du service et d'autres nécessitant des développements innovants qui peuvent être valorisés dans des journaux de méthodologie (du type JRSS B ou C).

D'un point de vue structurel, l'entretien avec les tutelles a permis au comité de comprendre que l'unité sera sous peu transformée en une équipe d'une unité de service de la Faculté de pharmacie de UPC. Ce regroupement s'inscrit dans une large dynamique de l'établissement pour diminuer le nombre d'unités. Cette intégration qui est subie par l'unité génère des réticences en son sein. Crainte de perdre en d'autonomie et crainte d'un changement de nature qui pourrait la rapprocher d'une plateforme de service. La tutelle, consciente des réticences de l'unité, tient un discours constructif, assurant au groupe qu'il pourra conserver son nom ainsi que sa dotation financière. De plus, elle souligne que l'équipe bénéficiera ainsi de la triple tutelle Paris Cité, Inserm et CNRS (au travers de la nouvelle unité) et de moyens inhérents (notamment pour la gestion). Dans cette dynamique, la nouvelle directrice de cette unité de service envisage de demander un poste d'IR Biostatisticien afin de venir aider le groupe sur la partie d'activité de service ou expertise de l'unité. Si ces conditions étaient remplies, l'unité aurait alors les moyens de mettre en œuvre son projet comme elle l'a fait au dernier mandat.

RECOMMANDATIONS À L'UNITÉ

Recommandations concernant le domaine 1 : Profil, ressources et organisation de l'unité

Le comité encourage l'unité à augmenter la visibilité de ses activités notamment au travers de la création d'un site web, et ce malgré le manque d'appui matériel de la part de la tutelle souligné par la direction. Cette visibilité est essentielle pour les doctorants et post-doctorants qui candidateront sur le marché du travail prochainement.

Recommandations concernant le domaine 2 : Attractivité

Augmenter la visibilité de l'unité ne peut qu'être bénéfique à son attractivité.

Recommandations concernant le domaine 3 : Production scientifique

Le comité encourage le groupe à poursuivre son activité de recherche multidisciplinaire avec le même dynamisme que celui de la période évaluée. Il l'encourage à préserver l'équilibre entre application et méthodologie et à renforcer la diffusion des méthodes innovantes auprès de la communauté de l'apprentissage statistique et de la modélisation, en publiant les contributions méthodologiques les plus significatives dans des journaux appropriés, et en participant à des conférences de cette communauté. Le comité encourage le groupe à mettre en valeur son travail et ses publications par exemple au travers d'un portail Hal dédié aux publications de l'unité, ou de l'équipe dans le contexte de la future fusion.

Recommandations concernant le domaine 4 : Inscription des activités de recherche dans la société

Le comité encourage l'unité à afficher de façon plus claire les liens entre ses activités et le projet Covid-19 Museum, qui concentre l'essentiel des efforts vers le grand public. Au-delà de ce projet, ciblé sur une période très particulière, la participation de membres de l'équipe à des actions de communications vers le grand public permettrait de valoriser les contributions les plus impactantes de l'unité.

DÉROULEMENT DES ENTRETIENS

DATE

Début : 18 janvier 2024 à 09h45

Fin : 18 janvier 2024 à 15h00

Entretiens réalisés : en présentiel dans les locaux de l'unité (Observatoire – Pharmacie, 4 avenue de l'Observatoire, 75006 Paris).

PROGRAMME DES ENTRETIENS

09:45 - Accueil du comité d'experts [public]

10:00 - Présentation de l'unité (bilan et trajectoire) et de son périmètre de recherche, 35mn + 5mn de questions [public]

10:40 - Exposés scientifiques (basées sur le portfolio), deux exposés de 8mn + 4mn questions [public]

11:00 - Visite des locaux

11:20 - Entretien avec les membres de l'unité (sans l'équipe de direction), 20mn

11:40 - Entretien avec les personnels scientifiques de rang B, 10mn

11:50 - Entretien avec les doctorants et postdocs, 15mn

12:20 - Déjeuner du comité à huis clos

13:45 - Entretien avec la direction, 30mn

14:15 - Huis clos du comité, 15mn

14:30 - Entretien avec la tutelle (Université Paris Cité), 20mn

14:50 - Huis clos final du comité, 10mn

15:00 - Fin de la journée

OBSERVATIONS GÉNÉRALES DES TUTELLES

Le Président

Paris, le 28 mars 2024

HCERES
2 rue Albert Einstein
75013 Paris

Objet : Rapport d'évaluation de l'unité DER-PUR250024161 - BioSTM - Biostatistique, traitement et modélisation des données biologiques.

Madame, Monsieur,

L'université Paris Cité (UPCité) a pris connaissance du rapport d'évaluation de l'Unité de Recherche **BioSTM - Biostatistique, traitement et modélisation des données biologiques**.

Ce rapport a été lu par la direction de l'unité, laquelle n'a pas noté d'erreurs factuelles, par le vice-doyen Recherche et le doyen de la Faculté de Santé d'UPCité, par la vice-présidente Recherche d'UPCité et par moi-même. L'ensemble des acteurs remercie le comité pour son travail d'évaluation.

Présidence

Référence

Pr/DGDRIVE/2023

Affaire suivie par

Christine Debydeal -
DGDRIVE

Adresse

85 boulevard St-Germain
75006 - Paris

www.u-paris.fr

Le doyen de la Faculté de Santé et moi même souhaitons souligner que l'unité BioSTM, créée en 2019, est une unité de recherche propre de l'université, focalisée sur la modélisation mathématique et l'apprentissage statistique appliquée aux données de santé. L'unité s'appuie sur l'analyse de données expérimentales d'autres équipes pour développer une méthodologie mathématique appliquée à la santé en regroupant des savoir-faire complémentaires. L'unité a su trouver sa place au sein de la faculté de pharmacie en développant un fort réseau collaboratif avec les autres UMR. Dans le futur, en accord avec UPCité et la faculté de santé, BioSTM intégrera l'unité mixte de service de l'UFR de pharmacie avec un statut particulier puisqu'elle conservera sa dotation de fonctionnement récurrent, ce qui n'est normalement pas le cas pour les équipes d'une UMS. Ce regroupement augmentera la visibilité et la cohérence institutionnelle de la future équipe qui sera intégrée dans une UMS collabellisée par UPCité, l'INSERM et le CNRS.

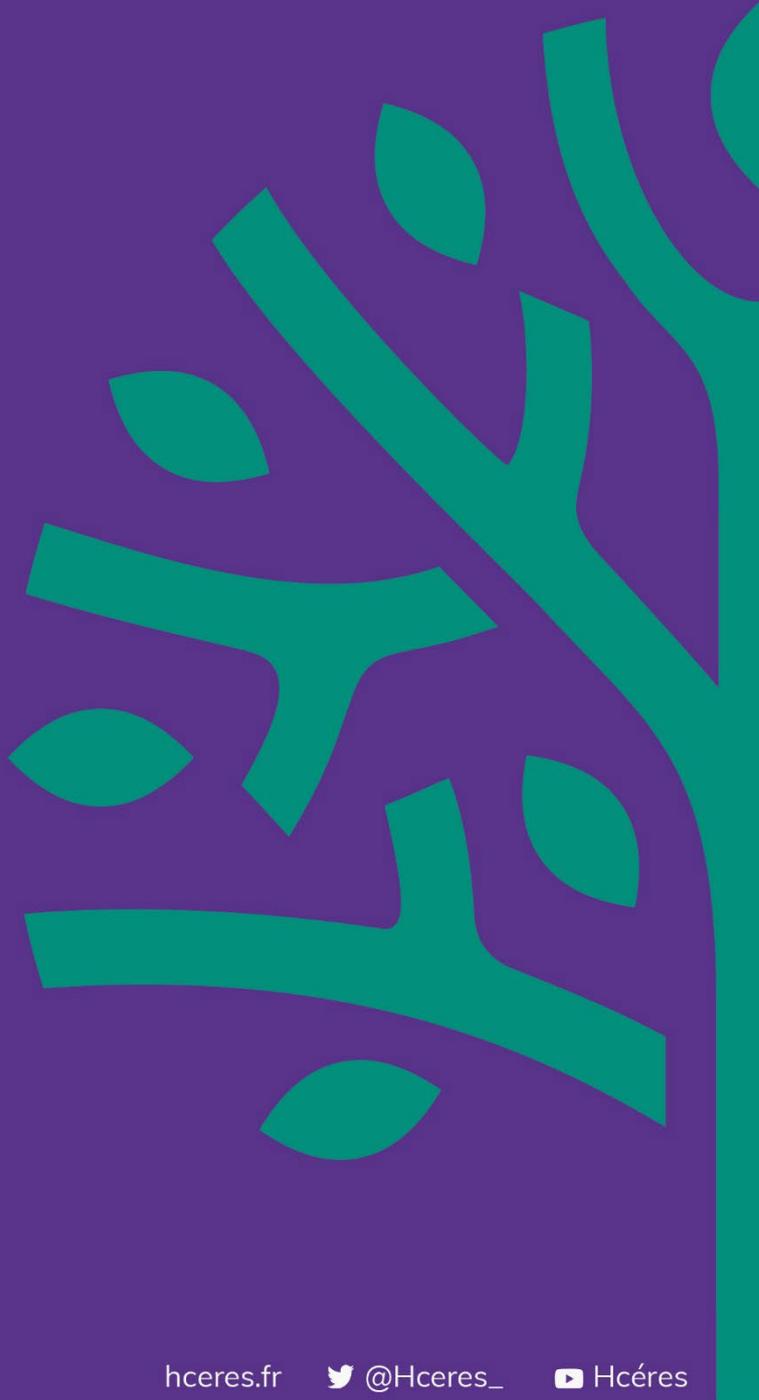
Je vous prie d'agréer, Madame, Monsieur, l'expression de ma considération distinguée.

Édouard Kaminski



Les rapports d'évaluation du Hcéres
sont consultables en ligne : www.hceres.fr

Évaluation des universités et des écoles
Évaluation des unités de recherche
Évaluation des formations
Évaluation des organismes nationaux de recherche
Évaluation et accréditation internationales



2 rue Albert Einstein
75013 Paris, France
T.33 (0)1 55 55 60 10

hceres.fr

 [@Hceres_](https://twitter.com/Hceres_)

 [Hcéres](https://www.youtube.com/Hceres)