Summary of Hospinnomics' annual workshop
Smart Data, Smart Methods: Challenges for Health Economics
September 28th – 29th, 2015

The workshop was organized around three panels (Monday 28th afternoon) and an interactive session on three different topics (Tuesday 29th morning).

Pierre Corvol (Chair, Hospinnomics Scientific Advisory Board) and **Pierre-Yves Geoffard** (PSE) opened Hospinnomics' annual workshop by discussing the opportunities tied to 'smart data'. More diverse data sources are now available and the possibilities to link different databases make them much more useful to develop knowledge and appropriate policies in health.

Panel 1: Smart data: impact on health care research and decision-making
Chair: **Pierre-Yves Geoffard** (PSE)

Randall Ellis (Boston University) gave examples of the opportunities opened by big datasets for healthcare research: statistical power is improved, studies on rare diseases become feasible, high-dimensional fixed effects for patients and doctors allow to control for the endogeneity of insurance plan choices, etc. Yet, many challenges remain for big data to be useful to research. Cleaning –namely avoiding duplicates or missing information– and organizing –shaping the data, constructing variables, defining weights for identical observations to reduce the size of the data– building such datasets represent an important investment. Making the data linkable to other sources implies a minimum of standardization but greatly multiplies the value of data. Clinical data, for example from x-rays or heartbeat monitors, constitute one of these new sources still too rarely analyzed, however promising for research.

Emmanuel Bacry (Ecole Polytechnique) addressed recent improvements in data science,which seek to analyze huge volumes of varied information. Unsupervised machine learning techniques are already extensively used in everyday life (internet searches, targeted advertising…) and academics could benefit from developing interdisciplinary partnerships with both public and private institutions in this area. For instance, Ecole Polytechnique has engaged in a partnership with CNAMTS to design machine learning algorithms on the French national database (SNIIRAM) in order to identify healthcare pathways in given pathologies, detect fraud or anomalies of pharmaceutical molecules. In this case, the database needs to be converted into a non-relational structure ("noSQL").

Geneviève Chêne (Univ. Bordeaux, AVIESAN) described smart data as "the DNA of public health research". However, access to data can remain a challenge given the existing restrictions on their storage, linkage or use. Unique identifiers are often needed and individual level data may threaten anonymity. Institutions like INSERM offer infrastructures to facilitate access to data and can play a key role to encourage multidisciplinary research while warranting data safety.

Franck Le Ouay (Honestica) presented his project of a patient's file that centralizes information on prescriptions, consultations, consumptions as well as results (DNA, blood tests…). This database can be used by healthcare professionals in order to make better informed decisions. Such a project raises the issue about how to encourage people to use reporting tools so that it generates useful data.

Panel 2: Smart methods for health economics
Chair: **Alberto Holly** (HEC Lausanne, Univ. Nova de Lisboa)

Alberto Holly (HEC Lausanne, Univ. Nova de Lisboa) introduced this panel by presenting one of his studies on asymmetric information and health insurance in Switzerland: the choice of a given insurance program reveals the person's risk type. Furthermore, moral hazard may arise as individuals decide to behave in a less healthy manner because they rely on their health insurance to cover the resulting costs.

Sherry Glied (NYU) began by giving an overview of smart evaluation methods for health economics. One of the promising research methods is using quasi experiments and available data to lower the cost of such evaluations. Examples included computer failure in some States on the day of a policy implementation, or the random allocation made to foster care assignment by social workers with different preferences. She encouraged the French health economists to take advantage of the many discontinuities that must exist in the French system to apply regression discontinuity methods. Her conclusion, however was that in some cases, what is missing is neither smart data nor smart methods, but rather smart research questions.

Andrew Jones (York University) presented how econometrics can offer useful tools to go beyond the linear effect assumption. He encouraged researchers to test different specifications and to look, not only at the effect on the mean, but on the whole distribution. 29% of healthcare costs are driven by only 4% of the patients: analyzing the tails of the distribution is often a most important issue.

Marc Gurgand (PSE, J-PAL) presented randomized controlled trials (RCTs), a method inspired from clinical trials but that is only beginning to be applied to the evaluation of non-pharmaceutical healthcare interventions. This method is especially useful to get influential evaluation results when there is no quasi experiment available. For example, it has been shown, by using random assignment of people to groups with more or less information about their past health spending, that this information helped them choose a more appropriate health insurance scheme. This method allows to avoid selection biases, although some limitations remain: attrition, difficulties to attain sufficient statistical power, or the costs associated with carrying out RCTs.

Antoine Bozio (PSE, IPP) invited the audience to consider an old, yet still very smart, method to evaluate public policies ex-ante: dynamic micro-simulation. While researchers have overlooked this tool, administrations still use it extensively to make predictions about long-term impacts of policies. Based on detailed information about the population's characteristics and on different equations to model behavior, micro-simulation can help anticipate future tendencies and policies' effectiveness with a view to allocate public funds more effectively.

Matteo Galizzi (LSE) explained how behavioral economics departs from some neoclassical assumptions relateed to psychological or sociological questions. This approach can be highly relevant when addressing health issues, and supports the need to develop data linkage methods between existing datasets and survey data. He also discussed some of the difficulties susceptible to arise with respect to RCTs: they cannot be implemented for policies with negative expected effects, they can lack external validity (compliance is usually higher during a trial than outside), difficulties can be encountered with the control group (if participants become aware of not receiving the treatment and of being still required to provide the data), with important associated costs.

Panel 3: Smart data, smart methods: quels impacts en pratique ? (Panel en français)
Chair: **Lise Rochaix** (PSE, Hospinnomics)

Jérôme Wittwer (Univ. Bordeaux, ISPED) a présenté l'évaluation de l'expérimentation des Territoires Santé Numérique (TSN). Dans cinq territoires (Aquitaine, Bourgogne, Rhône-Alpes, Ile-de-France, et La Réunion), des outils numériques ont été mis en place (hors télémédecine ou objets connectés) comme des portails d'information ou de collecte de données pour les patients, ou des plateformes de coordination des professionnels de santé, pour contribuer à réorganiser les soins

primaires. L'évaluation a pour objectif de montrer si ces outils sont effectivement utilisés et s'ils sont efficaces et efficients dans le parcours de soins (amélioration de la prise en charge, réduction des consommations de soins et des hospitalisations non pertinentes). Des territoires contrôles doivent être construits pour obtenir un contrefactuel avec une population et une offre de soins similaires à celles des territoires expérimentaux. Cet exercice est cependant très délicat et l'hypothèse d'une évolution commune dans l'organisation des soins sur les territoires créés sera difficile à vérifier. Si l'évaluation a vocation à s'appuyer principalement sur des données médico-administratives issues du SNIIRAM et des données de connexion, y seront également ajoutées des enquêtes auprès des patients et des professionnels pour vérifier l'acceptabilité de ces outils.

Marine Jeantet (CNAMTS, AT-MP) a présenté l'évaluation d'un programme d'accompagnement des victimes d'un accident du travail (AT) grave qui ne représentent que 2% de la sinistralité mais engendrent près de 40% des prestations de la branche. Un service de prévention secondaire est expérimenté dans cinq caisses primaires situées dans trois régions différentes. Un conseiller y prend en charge les patients depuis leur déclaration d'AT jusqu'à leur réinsertion professionnelle. L'objectif de l'évaluation, conduite par Hospinnomics, est d'identifier l'impact du programme sur le retour au travail, le niveau de séquelles, les indemnités journalières et les rentes, tout en le comparant au coût de mise en œuvre du service. Pour réconcilier le temps de la décision avec celui de l'évaluation, deux études sont mises en place : la première repose sur des témoins rétrospectifs appariés et permettra de donner des résultats dès 2017 ; la deuxième étude bénéficie d'une cohorte prospective obtenue par randomisation, mais qui ne pourra donner des résultats qu'en 2021. Pour que le principe de tirage au sort soit accepté par les acteurs, il a été décidé d'effectuer l'assignation au groupe témoin ou bénéficiaire avant la prise de contact avec l'assuré.

Axelle Charpentier (MAFEJ, DJEPVA) a apporté un retour d'expérience du Fonds d'expérimentation pour la jeunesse sur l'évaluation de programmes visant la réinsertion des jeunes sous main de justice. Les protocoles randomisés dans ce cadre n'ont pas été concluants à cause d'effectifs insuffisants, de tirage au sort parfois non respecté par les acteurs de terrain, ainsi que des difficultés dans le suivi de ces jeunes oscillant entre prison et milieu ouvert dans une temporalité difficile à prévoir. Ces projets montrent l'importance de l'appariement entre porteur de projet et évaluateur pour la définition d'une méthodologie qui corresponde aux réalités de terrain et une sensibilisation suffisante des opérateurs (tant les acteurs locaux que ceux au niveau politique supérieur qui assurent le financement) au respect du protocole d'évaluation. Les études d'impact quantitatives ont généralement un coût par personne plus faible que les enquêtes qualitatives reposant sur des entretiens.

Franck von Lennep (DREES) a parlé des données de santé existantes, ainsi que des évolutions à venir dans ce domaine, notamment avec l'article 47 du projet de loi de santé. L'exhaustivité des données médico-administratives issues du SNIIRAM est un grand avantage par rapport à des données d'enquête. Cependant, leur appariement avec d'autres sources de données permettrait de les rendre plus « smart » et donc plus utiles à la décision. Par exemple, en incluant les informations des organismes complémentaires, les analyses des restes à charge seraient basées sur les véritables montants plutôt que sur des simulations. De même pour le revenu ou la catégorie socio-professionnelle, qui seraient de meilleurs indicateurs du niveau économique et social de chaque individu que l'indicateur reposant sur une cartographie des zones de précarité. L'ajout de variables plus médicalisées (résultats d'examens médicaux) ou des causes de décès permettrait des analyses plus pertinentes sur l'état de santé de la population. Néanmoins, la protection des données personnelles reste une priorité, ce qui implique notamment l'interdiction de tout usage commercial de ces données.

Namik Tarik (AP-HP, DOMU) a évoqué la richesse des bases de données détenues par les établissements de santé, et l'AP-HP en particulier. Ces données sont cependant trop rarement exploitées dans des optiques autres qu'administratives. Le croisement de données du PMSI avec les prescriptions pharmaceutiques et les dossiers médicaux des patients serait possible pour répondre à des questions comme la pertinence de la prescription des médicaments sur la liste en sus, la description des événements précédant un décès, ou encore les infections sur les sites opératoires. Ceci suppose cependant de développer des techniques pour analyser les nombreux champs de texte libre qui font la richesse de ces informations.

Marie Zins (Univ. de Versailles Saint-Quentin-en-Yvelines) a décrit la cohorte épidémiologique Constances, qui suivra 200.000 adultes représentatifs de la population générale en s'appuyant sur le réseau des centres de santé. L'objectif est d'étudier l'évolution de la santé de la population française, en s'intéressant de près aux causes des maladies. Un examen clinique important à l'inclusion, et qui comporte notamment des données biologiques, permet le recueil de nombreuses données de base. Le suivi se fait ensuite annuellement de manière plus légère, en particulier grâce aux extractions des bases de données administratives : CépiDC (causes de décès), SNIIRAM (consommations de soins) et base CARRIERES de la CNAV (parcours professionnel). Des données sont aussi recueillies sur l'environnement dans lequel évoluent les participants (pollution, entreprises, proximité des médecins). Le consentement des personnes est un enjeu majeur dans ce cadre et la CNIL a ainsi exigé qu'il porte également sur l'appariement avec les données administratives. La construction d'une cohorte de cette envergure s'avère très coûteuse – ici financée par les investissements d'avenir – mais elle a vocation à être utilisée par de nombreux chercheurs, à la fois par la mise à disposition de la base de données et par la possibilité d'accéder aux volontaires pour des enquêtes ciblées (exemple d'une enquête de la DREES sur les délais d'attente pour les rendez-vous médicaux). D'après l'expérience de la cohorte GAZEL, le taux de réponse au questionnaire annuel devrait rester élevé, car les personnes apprécient le fait qu'on s'intéresse à elles. La représentativité de la cohorte est par ailleurs assurée grâce à des repondérations s'appuyant sur un échantillon de 400.000 non-participants suivis dans les bases administratives.

Cynthia Fleury (Chaire Philosophie, AP-HP) a apporté une réflexion sur les enjeux éthiques liés à l'analyse des données de santé. Elle a rappelé les principes de protection des données personnelles fixés par la Loi « Informatique et libertés » : finalité, proportionnalité, pertinence, durée limitée de conservation, sécurité et confidentialité, transparence, respect du droit des personnes. Ceci fait écho au droit à l'auto-détermination informationnelle d'une personne qui lui permet de décider de la communication et de l'utilisation de ses données personnelles. En incluant cette dimension dès le début des interventions et projets de recherche, on permet aux personnes de ne pas se sentir forcées dans le partage de leurs données mais au contraire de pouvoir choisir de participer à l'élaboration de nouvelles connaissances. Par ailleurs, si du point de vue du chercheur la légitimité du travail sur les données est proportionnel à l'utilité de ses analyses, en termes de production de connaissances, cette légitimité est différente de celle du politique, qui est lui un représentant de la nation. Il est donc essentiel de concilier les deux légitimités, par exemple en misant sur une acculturation du politique à la science et à l'utilisation des résultats de recherche dans la décision.

Working group 1: Inequalities
Chair: **Florence Jusot** (Univ. Dauphine, Chaire Santé)

This working group focused on inequalities in healthcare for different population subgroups, echoing Andrew Jones' advice during panel 2 to analyze distributions and not only mean effects.

distributional considerations are particularly relevant when it comes to out-of-pocket expenses and three ongoing research projects were discussed. Within a study carried out by Hospinnomics for AP-HP, **Bénédicte Apouey,** PSE, Hospinnomics' affiliate, is using data from 37 AP-HP hospitals to study the distribution of out-of-pocket expenses for patients. **Grégoire de Lagasnerie** from CNAMTS works on a sample of beneficiaries extracted from the SNIIRAM ("l'échantillon généraliste des bénéficiaires" – EGB) to study the redistribution that takes place, due to the fact that out-of-pocket expenses vary with patients' age, health status or income. While out-of-pocket expenses are highly concentrated, there is still a redistribution over the life cycle and from the healthier to the unhealthier for whom out-of-pocket payments represent a smaller proportion of total health costs. However, this administrative database has important limitations: health status is approximated with information on long term conditions ("affection de longue durée" – ALD) or algorithms using actual medical care consumption; a deprivation index based on patients' living area provides a proxy for socio-economic status, in the absence of information on actual; last but not least, complementary insurance reimbursements are not known, meaning that out-of-pocket expenses correspond to simulated rather than to actual values. The MONACO project (« Méthodes, outils et normes pour la mise en commun

des données des assurances complémentaire et obligatoire »), in which **Paul Dourgnon** from IRDES is involved, tries to circumvent these issues by merging data from mandatory health insurance, complementary health insurance plans and data from a national health survey ("enquête sur la santé et la protection sociale – ESPS).

Access to healthcare can be subject to inequalities depending on migration status. **Dr Claire Georges-Tarragano** is in charge of the PASS program ("Permanences d'Accès aux Soins de Santé" – PASS) at Saint-Louis AP-HP hospital in Paris. This outpatient care unit is dedicated to vulnerable populations (deprived patients, undocumented migrants…). Their medico-social situation can be highly complex, so that a multidisciplinary team is necessary to deal with these issues (including lack of health insurance or linguistic barriers). No other source of data can be found on undocumented migrants' health in France, which makes this clinical care setting especially interesting for researchers. One of its positive impacts for the social health care system is the rationalization of emergency department visits for these patients. Nevertheless, a complete evaluation of this program and of other migrants-related health policy is necessary in order to assess and to improve their efficiency.

Working group 2: Evaluation of health care interventions
Chair: **Sandy Tubeuf** (Leeds University)

This working group started with a presentation of a therapeutic education intervention for patients with renal failure. **Jean-Claude K. Dupont** from Hospinnomics explained that Saint-Louis hospital in Paris is developing an internet platform with patient information on the disease as well as social media features allowing patients to interact with medical staff. A randomized experiment will be conducted in partnership between Hospinnomics and J-PAL Europe (PSE) to identify the mean effect from access to this dedicated platform, and the added value of social media features, on patients' quality of life and intermediary clinical outcomes (e.g. reduced adverse events).

The discussion then focused on patients' involvement in research. Jean-Claude K. Dupont explained the importance of developing partnership approaches as in community-based participatory research in the health sector, for example by involving patients and patients' association from the inception of research projects. **Marine Genton**, (Lyon I) working on individuals' perceptions of the risk of cancer, highlighted the importance of conducting qualitative studies with patients before building a quantitative questionnaire. Indeed, focus groups may reveal that some concepts, held as self-evident for researchers, have a very different meaning for patients.

Lise Rochaix underlined the fact that preferences are very dependent on which perspective has been adopted for the evaluation. For instance, **Christine Leclainche**, from ENS Cachan, is studying the difference between patients' preferences and general population's preferences about cancer screening, an issue which is especially relevant in France where the rate of mammography is lower than in other European countries. **Ivy Lu** from Hospinnomics pointed out that patients are not the only stakeholders whose preference can impact on health care interventions. Physicians for example may also differ in their willingness to prescribe breast cancer screening. healthcare professionals' perspective should also be taken into account in evaluation. Expertise in behavioral economics is clearly central for the evaluation of health care interventions.

Working group 3: Drug regulation
Chair: **Margaret Kyle** (MinesTech)

The third working group focused on drug regulation in different countries. Regulation in this area is based on the rationale that drugs are no ordinary products; their use needs to be supervised in order to promote public health. Drug regulation includes the following activities: licensing and monitoring of manufacturers and distribution intermediaries, assessment and monitoring of medicines' safety and efficacy, market authorization. These tasks are undertaken by national regulatory agencies, but also by more global authorities – for example at the European level. **Margaret Kyle** stressed the fact that this sector is subject to free-riding as some institutions rely on others to do the costly reviewing process of

drugs. Moreover, the certification process of a drug's cost-effectiveness is complex and costly for manufacturers because agencies have different requirements. One way to reduce the cost of medicines could be to standardize this process across countries, which is not incompatible with the freedom of national agencies to fix different prices.

Drug regulation can impact on the development of innovative medicines. **Philippe Gorry** (Bordeaux University) is leading a project funded by INCa on innovation and inequality in the field of rare cancers drugs. **Eliana Barrenho** (post-doc at Hospinnomics in 2015), who is also involved in this project, has been studying R&D in drugs at an international scale by using various data sources. As pointed out, one difficulty of working with different sources of data lies in the differences in disease coding systems. All in all, these projects help to address the questions of the financial sustainability of innovation and of the incentives to innovation set up by the regulators.

The process of drug registration can be rather long and complicated. In France, it can last up to one and a half year between European agreement and marketing authorization and drug pricing at the national level. The economic implications of this system and more specifically of the regulation called ATU (authorization for temporary use) are studied by **Ivy Lu** (post-doc at Hospinnomics in 2015).

Regulation has a strong impact on the pricing of drugs. Analyzing price-sensitivity of doctors or patients is not trivial, and can have strong impacts on public health. Research can be hindered by the secrecy that surrounds the actual prices of drugs: in France, for example, hospital pharmacies negotiate prices under the cap defined by Social Security. Through a partnership with AGEPS (*Agence Générale des Equipements et Produits de Santé*), Hospinnomics plans to conduct research on the topic of drug prescriptions at hospital level.

Martin Hirsch (AP-HP) concluded Hospinnomics' annual workshop by thanking the members of the Scientific Advisory Committee, and also Lise Rochaix for all the work accomplished in one year. The chair fits perfectly into AP-HP's new strategy: Openness, by collaborating with multiple partners; Unity, by connecting different databases and making them available for research; Innovation, by introducing economics in the hospital and offering evaluation tools to guide medical and management practices. He also suggested "time" as a suitable topic for future research, in order to strengthen the interactions between economics, philosophy and epidemiology. Finally, considering that the shortage is on smart research questions more than on smart data, as suggested by Sherry Glied, Martin Hirsch invited the entire community to ask questions on which economics can provide answers, not only with a view to "save costs", but as a true science interacting with other sciences in the health care sector.