

# Le défi de la recherche sur les soins en Suisse avec des données de routine

Claudia Berlin,  
Radoslaw Panczak,  
Matthias Egger

Institut de médecine sociale et préventive, Université de Berne

Les données médicales de routine et la recherche sur les soins qui en découle constituent un élément majeur de la prise de décision en politique de santé. Les études sur les différences régionales en matière de soins jouent un rôle important. L'objectif consiste donc à les comprendre et, lors de variations injustifiées, à mettre en route des processus de réformes pour les réduire. Pourquoi, par exemple, selon la *statistique médicale des hôpitaux*, le canton de Berne fait état de deux fois plus de cathétérismes cardiaques que le canton de Vaud (fig. 1) [1]? S'agit-il d'une surabondance de l'offre dans le canton de Berne ou d'une insuffisance dans le canton de Vaud? Dans ce dernier, les examens ont-ils plus souvent lieu en ambulatoire? Le besoin est-il moins élevé dans le canton de Vaud en raison de l'état de santé de la population, ou la demande est-elle induite par l'offre du canton de Berne? Seules des analyses approfondies permettent de répondre à ces questions si les données de routine sont réparties selon des petites unités géographiques et qu'il est possible de les recouper entre elles.

Dans un manifeste publié l'an dernier, Santé publique Suisse a exigé des «*données de meilleure qualité pour augmenter l'efficacité du système de santé*» [2]: l'exhaustivité et la disponibilité mais aussi l'appariement et la compatibilité des données ne sont pas satisfaisants. Nous vous présentons ici les principales sources de données, et relatons nos expériences concernant leur utilisation dans le cadre de la recherche sur les soins.

## Appariement de banques de données

Le couplage de séries de données peut être réalisé grâce à un numéro d'identification explicite (par exemple un numéro de personne comme en Suède) ou par un chiffre établissant un lien entre plusieurs variables (par ex. date de naissance, lieu de domicile, sexe, état civil, nationalité pour relier les données de recensement et les données de mortalité du Swiss National Cohort [3]). Si le couplage ne peut pas se faire au niveau de l'individu, les données peuvent être reliées entre elles selon des informations géographiques. La prévalence d'une maladie peut ainsi être établie par commune en associant les données sur la fréquence de la maladie à la statistique de la population. Cet appariement devient compliqué si les séries de données ont été créées selon des références géographiques différentes, qui ne sont pas ou qu'en partie compatibles entre elles (par ex. différence entre

**En santé, une des priorités de la recherche réside dans l'identification et l'analyse des différences régionales en matière de prise en charge médicale. Une recherche sur les soins basée sur des preuves requiert cependant des données de routine susceptibles de répondre aux exigences concernant l'exhaustivité des données, leur disponibilité, appariement et compatibilité. Ces exigences ne sont qu'en partie remplies par les données médicales de routine existantes («Des données de meilleure qualité pour augmenter l'efficacité du système de santé», Santé publique Suisse). Dans cet article, nous abordons les enjeux en lien avec quelques-unes des principales sources de données et les défis que nous avons eu à relever lors du couplage de différentes séries de données et de leur analyse.**

code postal et commune). Dans ce cas, il faut choisir une unité géographique plus grande, par ex. le canton (fig. 1). Mais de manière générale, il est préférable de prendre la plus petite unité possible car une analyse selon des unités plus grandes peut toujours être réalisée ultérieurement.

Les difficultés s'accroissent si l'analyse des données se fait dans le cadre d'une longue série chronologique. En effet, la zone géographique correspondant à un code postal ainsi que la superficie, le nom ou le numéro d'une commune peuvent changer avec le temps, les communes pouvant fusionner, être divisées ou rattachées à une autre entité. Lorsque ce type de modifications apparaît dans plusieurs séries de données établies selon différentes unités de référence (géographiques ou temporelles), l'appariement des données devient alors compliqué et il est difficile de mettre en place un modèle stable pour l'analyse géographique.

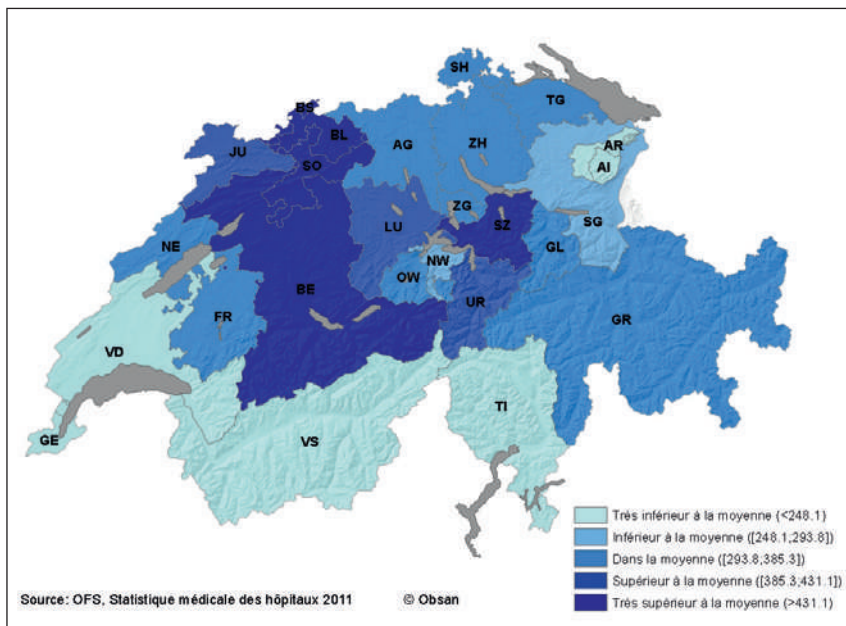
## Séries de données du secteur hospitalier

Tous les hôpitaux suisses ont l'obligation de récolter des données sur les patients traités ainsi que sur leurs ressources humaines et leurs infrastructures techniques. Les séries de données recueillies font l'objet d'une première préparation par l'*Office fédéral de la*

Correspondance:  
Prof. Dr Matthias Egger  
Institut de médecine sociale et préventive (ISPM)  
Université de Berne  
Finkenhubelweg 11  
CH-3012 Berne  
Tél. 031 631 35 11  
Fax 031 631 35 20  
egger[at]ispm.unibe.ch

Figure 1

Taux standardisé de cathétérismes cardiaques par canton en 2011 (pour 100 000 habitants).



statistique (OFS) qui les met ensuite à disposition pour l'analyse des soins hospitaliers sous la forme de :

- la statistique des hôpitaux,
- la statistique médicale des hôpitaux.

La *statistique médicale des hôpitaux* (MEDH) répertorie des informations concernant tous les cas hospitaliers traités en stationnaire en Suisse. Un nombre important de variables est collecté pour chaque cas (tabl. 1). La *statistique des hôpitaux* (STH) recense des données relatives à l'infrastructure, dont notamment le nombre de lits, le nombre d'appareils de tomographie par résonance magnétique ou le nombre d'employés. Ces deux statistiques peuvent être reliées entre elles via le numéro d'identité de l'hôpital. Par ailleurs, 706 régions appelées régions *MedStat* ont été introduites dans le but de garantir l'anonymat des patients et de codifier les indications géographiques concernant le patient (lieu de domicile) et l'hôpital (lieu du site). Comme une région *MedStat* regroupe plusieurs codes postaux et que ceux-ci changent fréquemment, l'OFS publie chaque année une nouvelle table pour leur transcodage.

#### Patients zurichois dans les hôpitaux schaffhousois

Dans différents projets de notre groupe de recherche, nous utilisons les deux statistiques ci-dessus de l'OFS des années 2008 à 2011 dans le but d'analyser la prise en charge hospitalière de manière générale mais aussi selon les maladies. Nous nous basons sur les régions *MedStat* pour la localisation géographique et pour la représentation des structures de soins. Mais

les vérifications de la plausibilité de nos données nous a rapidement intrigués. En effet, la comparaison des nombres de cas en fonction des régions a mis en évidence d'énormes écarts d'une année à l'autre, ainsi que des trajectoires de patients très inhabituelles. A titre d'exemple, citons que la moitié des patients habitants dans la région *MedStat* ZH52 (à l'est du lac de Zurich) en 2008 se sont faits soigner dans les hôpitaux schaffhousois alors qu'ils n'étaient plus que 2 en 2011.

A l'origine de ces fluctuations préside un redécoupage des régions *MedStat* entre 2007 et 2008, réalisé dans le but d'adapter les régions *MedStat* aux régions MS (mobilité spatiale) et aux bassins d'emploi pour des raisons de compatibilité. Plusieurs dénominations des régions ont ainsi été conservées alors que la zone géographique qu'elles recouvrent est sensiblement différente. Pour revenir à notre exemple, la région ZH52 recouvre en 2007 la région au sud de la frontière avec le canton de Schaffhouse et, à partir de 2008, cette même dénomination correspond à la région à l'est du lac de Zurich. D'autres recherches nous ont permis de nous apercevoir qu'un des fournisseurs de logiciels n'a transmis la mise à jour logicielle de la nouvelle table de transcodage qu'en 2009 ou 2010 aux hôpitaux. Par conséquent, les institutions concernées ont continué de coder avec l'ancienne table, affectant ainsi leurs patients à des régions obsolètes.

Tableau 1

Quelques-unes des variables recueillies lors de la collecte des données dans le domaine hospitalier stationnaire.

- Numéro d'identification du cas
- Numéro d'identification du patient
- Numéro d'identification de l'hôpital
- Sexe
- Age
- Région de domicile du patient
- Site de l'hôpital
- Jour de l'admission
- Jour de sortie
- Mode d'admission (par ex. urgence)
- Décision de sortie (par ex. décès, transfert)
- Diagnostic principal
- Diagnostic supplémentaire
- Code CHOP*
- APDRG (SwissDRG depuis 2012)**

\* La CHOP (classification suisse des interventions chirurgicales) sert à coder les traitements hospitaliers.

\*\* Le système de forfaits par cas APDRG (All Patient Diagnosis-Related Groups), appliqué dans quelques cantons pour la rémunération des prestations des hôpitaux de soins somatiques aigus, a été remplacé par le système SwissDRG introduit en 2012 dans toute la Suisse.

### **Davantage de patients hospitalisés que d'habitants**

Pourquoi le nombre d'hospitalisations dans la région MedStat à l'est de Winterthour dépasse-t-il le nombre de personnes domiciliées dans cette région? Il s'est avéré que le code postal utilisé pour répartir les patients dans cette région ne correspondait pas à leur code postal réel. En effet, il n'est pas rare qu'une commune s'étende sur plusieurs codes postaux ou inversement qu'un code postal soit réparti sur plusieurs communes. Ainsi, l'affectation des patients de Winterthour s'est faite sur la base du code postal de Winterthour, qui est également celui d'une commune à l'est de la ville. Dans un premier temps, les patients ont donc été attribués correctement à la ville de Winterthour. Mais dans un deuxième temps, comme le code postal utilisé correspondait également à celui d'une commune située en dehors de la ville de Winterthour, la table de transcodage a attribué les patients de ce code postal à la région MedStat à l'est de Winterthour.

### **Disparition d'hôpitaux et autres problèmes de répartition**

La fusion d'hôpitaux se traduit par la fusion des données des hôpitaux concernés, qui sont transmises sous un seul et même numéro d'identification alors que le regroupement d'hôpitaux réunit des institutions installées sur des sites différents, et donc dans des régions MedStat différentes. Dans ce contexte, il est impossible d'attribuer correctement les patients à la région MedStat de l'hôpital.

Les patients hospitalisés plusieurs fois peuvent normalement être identifiés grâce à un numéro d'identification qui se compose du nom, prénom, date de naissance et sexe rendus anonymes par une fonction de hachage (Hash Code). Cependant, les numéros d'identification créés de cette manière ne sont pas suffisamment précis. Des problèmes apparaissent lorsque le patient change de nom car un nouveau numéro lui est attribué alors qu'il s'agit de la même personne.

### **Décédé deux fois**

La *statistique des institutions médico-sociales* (MEDSO) recense les données concernant toutes les institutions médico-sociales en Suisse. Dans le cadre d'un projet sur les soins en fin de vie, nous avons couplé les données de la MEDH et de la MEDSO des patients décédés en utilisant leur numéro d'identification anonyme. Nous avons alors constaté que le codage de certaines personnes correspondait dans les deux statistiques à celui de «décédé» bien que le décès ne peut avoir été prononcé que dans une des deux institutions. Cela s'explique en partie par des erreurs de codage et par des numéros d'identification qui ne sont pas explicites. Par ailleurs, les séjours simultanés dans différentes institutions créent également des problèmes, notamment lorsque les lits restent

libres jusqu'au retour des patients adressés à une autre institution. De ce fait, la durée réelle du traitement dans une institution ne se laisse pas déterminer de manière claire et précise.

### **Manque de netteté et erreur du lieu de domicile**

Pour la recherche sur les soins, il est impératif d'avoir des indications exactes concernant le nombre, la spécialisation et le recours à des fournisseurs de prestations médicales. Dans la *statistique des hôpitaux*, les établissements avec plusieurs types d'activité (p. ex. soins aigus, psychiatrie, gériatrie) ne mentionnent cependant que le nombre total de leurs lits, sans répartir ce chiffre par service spécialisé.

Dans le secteur ambulatoire, les numéros du Registre des codes-créanciers (numéros RCC) des médecins admis à pratiquer ne permettent pas toujours de remonter à un seul médecin. Il est également possible que tout un cabinet de groupe facture avec un seul et même numéro RCC.

Enfin, nous avons également constaté que la variable «domicile» était souvent fautive dans la *statistique médicale des hôpitaux* pour la période entre 2008 et 2010. Il est donc déconseillé de procéder à des analyses à petite échelle géographique avec les données des régions MedStat. Les analyses qui se fondent sur ces données (par ex. dans le cadre de la planification cantonale des soins) devraient par conséquent faire l'objet de la plus grande réserve.

### **Propositions de solutions**

Ce sont les fournisseurs de prestations ou les cantons qui attribuent les cas dans leur région MedStat respective. De ce fait, l'OFS n'a aucune influence sur cette étape et ne peut pas la vérifier. Il a cependant invité les hôpitaux et les cantons à utiliser la bonne table de transcodage pour attribuer leurs données à la région MedStat qui leur correspond. Depuis 2011, l'OFS tente également de maîtriser le problème soulevé par la fusion d'hôpitaux avec l'introduction d'une nouvelle variable «site hospitalier».

Nous soutenons l'exigence formulée par *Santé publique Suisse* de créer un numéro d'identification anonyme et uniforme [2], applicable à toutes les séries de données médicales dans le but d'améliorer le couplage et le recouplement anonyme des données. Une autre possibilité serait de coupler les informations grâce à des adresses géocodées, c'est-à-dire par l'affectation des coordonnées géographiques à une adresse, permettant une géolocalisation précise du lieu et facilitant l'attribution d'une unité administrative ou géographique aux données récoltées. Les données requises existent en Suisse (par ex. les coordonnées des bâtiments de l'*Office fédéral de topographie [Swisstopo]*). Il est bien évident que le géocodage des adresses des patients et des fournisseurs de prestations mais aussi leur archivage doivent répondre aux dispositions extrêmement strictes en matière de

protection des données. Cette tâche, et celle qui consisterait également à garantir l'uniformité des numéros d'identification, pourraient être confiées au *Trust Center* présenté dans le manifeste [2]. Si les informations géographiques d'un patient ou d'un hôpital sont nécessaires pour une analyse, le *Trust Center* peut simplement attribuer les coordonnées d'une adresse à une unité géographique (p. ex. une commune) et ensuite les transmettre sans avoir divulgué de coordonnées nominatives.

Les statistiques MEDH et STH utilisent la répartition géographique des régions Medstat qui seront également employées pour la collecte des données dans le secteur ambulatoire (projet MARS: *Modules*

*Ambulatoires des Relevés sur la Santé*). Nous proposons de ne pas donner suite à cette manière de procéder mais plutôt de géocoder les adresses, cette méthode permettant une nette amélioration en matière de fiabilité des variables géographiques et par conséquent d'analyse des données.

### Conclusion et perspective

Les exemples mentionnés dans ces colonnes ne se concentrent que sur une partie des séries de données susceptibles de revêtir de l'importance pour la recherche sur les soins. L'expérience montre cependant que les problèmes n'apparaissent qu'au moment de l'utilisation et de l'évaluation des données. Les données récoltées doivent régulièrement faire l'objet d'une analyse dans le but de les améliorer et ainsi de mieux étayer la réflexion précédant les décisions en politique de santé. Les problèmes soulevés pourraient être réglés de manière relativement facile. L'avenir nous dira quelle est la volonté politique en la matière. Mais la tournure prise dans le domaine des registres sur le cancer nous donne bon espoir qu'une recherche sur les soins fondée et largement soutenue sera possible en Suisse d'ici quelques années.

### Références

- 1 Pellegrini S, Kohler D, Otto S. Variations géographiques dans les soins de santé. La situation en Suisse. Un projet de l'OCDE. Obsan: Neuchâtel; 2014.
- 2 Public Health Schweiz. Manifest. Bessere Gesundheitsdaten für ein effizienteres Gesundheitswesen. Disponible (en allemand) sous: [www.hausarztmedizin.uzh.ch/aktuell/130816-ManifestGesundheitsdaten-D-def.pdf](http://www.hausarztmedizin.uzh.ch/aktuell/130816-ManifestGesundheitsdaten-D-def.pdf) (accessed 2 Nov 2014).
- 3 Bopp M, Spoerri A, Zwahlen M, Gutzwiller F, Paccaud F, Braun-Fahrlander C, et al. Cohort Profile: the Swiss National Cohort—a longitudinal study of 6.8 million people. *Int J Epidemiol.* 2009;38:379–84.

### Groupe de dialogue «Recherche prioritaire: recherche sur les soins»

Pour le corps médical, la recherche sur les soins constitue un domaine scientifique important porteur d'avenir. Dans un secteur de la santé en pleine mutation (nouveaux modèles de financement, évolutions démographiques, hausse des coûts de la santé, etc.), il est impératif que la recherche sur les soins bénéficie d'un ancrage académique. Afin de pouvoir créer les bases scientifiques d'une recherche indépendante dénuée de tout intérêt particulier, la Fédération des médecins suisses (FMH), la Conférence des sociétés cantonales de médecine (CCM) et NewIndex soutiennent le groupe de dialogue «Recherche prioritaire: recherche sur les soins» de l'Institut de médecine sociale et préventive de l'Université de Berne.

A l'instar d'un forum, ce groupe de dialogue a pour but de solliciter l'échange d'informations: les représentants des organisations mentionnées et des groupes de recherche discutent régulièrement des travaux en cours et des projets à venir dans le domaine de la recherche sur les soins. Par ailleurs, le groupe de dialogue vise à sensibiliser le corps médical à la recherche sur les soins et à en favoriser l'acceptation. Il est ouvert aux propositions en ce qui concerne les sujets à traiter, les questions, les discussions ou les demandes d'informations supplémentaires. La division Données, démographie et qualité (DDQ) de la FMH assure la coordination du groupe de dialogue et se tient à disposition pour tout complément d'information par courriel à [ddq\[at\]fmh.ch](mailto:ddq[at]fmh.ch) ou par téléphone au 031 359 11 11.