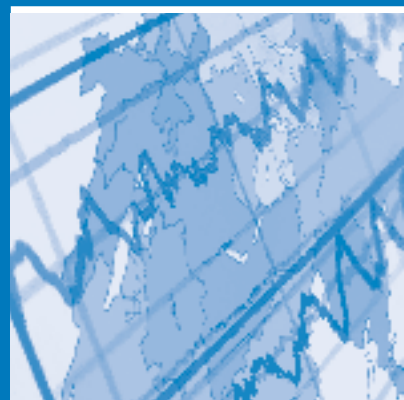


Rapports d'experts et de recherche pour l'assurance-maladie et accidents

# Analyse des prestations prises en charge par l'assurance maladie obligatoire des soins

Cadre conceptuel et étude de faisabilité centrée sur trois pathologies : cancer, diabète et affections mentales



Schweizerische Eidgenossenschaft  
Confédération suisse  
Confederazione Svizzera  
Confederaziun svizra

Département fédéral de l'intérieur DFI  
Office fédéral de la santé publique OFSP



Rapports d'experts et de recherche pour l'assurance-maladie et accidents

# **Analyse des prestations prises en charge par l'assurance maladie obligatoire des soins**

Cadre conceptuel et étude de faisabilité centrée sur trois pathologies : cancer, diabète et affections mentales

Yves Eggli, Patricia Halfon, Mehdi Chikhi, Lam Nguyen, Anne Decollogny, François Weissbaum

Septembre 2006

L'office fédéral de la santé publique publie dans sa série «Rapports d'experts et de recherche pour l'assurance-maladie et accidents» des analyses et des rapports de recherches sur des sujets d'actualité dans le domaine de l'assurance-maladie et accidents pour les rendre accessibles au grand public et encourager la discussion. Les analyses présentées par les auteurs ne reflètent pas forcément l'opinion de l'Office fédéral de la santé publique.

## **Impressum**

© Office fédéral de la santé publique (OFSP)

Editeur: Office fédéral de la santé publique

Date de publication: septembre 2007

Reproduction partielle autorisée, sauf à des fins commerciales, avec mention de la source et envoi d'un justificatif à l'OFSP, section Statistiques et mathématiques.

Informations supplémentaires: OFSP, Unité de direction Assurance maladie et accidents, Till Bandi, téléphone +41 (0)31 322 91 31  
till.band@bag.admin.ch, www.bag.admin.ch

Commande de la publication:

OFCL, Diffusion publications, CH-3003 Berne

www.bbl.admin.ch/fr/ Rubrique Produits et services – Commande en ligne

Numéro de commande OFCL: 316.720

Numéro de publication OFSP: BAG KUV 9.07 300 kombi 60EXT07002

Inprimé sur du papier blanchi sans chlore

# Table des matières

Avant-propos	1
Vorwort	2
Résumé	3
Zusammenfassung	6
<b>1</b> Cadre conceptuel	10
1.1 Introduction	10
1.2 Description des données	15
1.3 Anonymisation et protection des données	17
1.4 Méthode d'identification des maladies	20
1.5 Avertissement méthodologique général	23
<b>2</b> Etude de faisabilité	25
2.1 Indicateurs de couverture	25
2.2 Indicateurs de financement	27
2.3 Indicateurs de morbidité	28
2.4 Indicateurs d'accessibilité	29
2.5 Indicateurs de pertinence	30
2.6 Indicateurs d'adéquation	33
2.7 Indicateurs d'effectivité	36
2.8 Indicateurs d'efficacité	37
2.9 Indicateurs d'efficience	38
<b>3</b> Conclusions et recommandations	42
<b>4</b> Utilisation des données mises à disposition	46
A I Liste des données fournies par les assureurs maladie	46
A II Liste des données utilisées de la statistique médicale des hôpitaux	47
A III Médicaments marqueurs d'un cancer	47
A IV Médicaments marqueurs d'un diabète	54
A V Médicaments marqueurs d'une maladie mentale	56
E I Prise en charge d'un assuré diabétique de 63 ans	64
E II Prise en charge d'une assurée souffrant de problèmes psychiatriques	65
E III Prise en charge d'un assuré souffrant d'un cancer	66
Bibliographie	68
Protection des données Rapport du service de cryptologie de l'armée du DDPS	71



## Avant-propos

Depuis des années, le système de santé suisse est considéré comme statistiquement lacunaire avec peu d'espoir d'amélioration à court terme dans ce domaine en constante évolution.

A ce titre, le potentiel inexploité des données administratives des assureurs-maladie permettraient facilement de développer différentes analyses plus pertinentes que celles réalisées jusqu'ici. Elles pourraient aussi être exploitées avec peu d'effort pour améliorer cette situation insatisfaisante. Il est évident que les bases de données suivantes pourraient être exploitées :

- Dans le cadre des processus administratifs, les assureurs saisissent les données concernant les personnes assurées (âge, sexe, etc.) et les prestations fournies (notamment le montant des factures, la participation aux coûts, le genre de prestations, les fournisseurs de prestations);
- La statistique médicale des hôpitaux livre des données détaillées concernant les patients traités avec les diagnostics et les interventions;
- L'organe de facturation des médicaments fournit les indications concernant les médicaments et le pharmacode qui sont remis aux patients.

Ces données réunies permettent à partir des traitements médicaux et pharmaceutiques d'en déduire des informations supplémentaires. Une condition est cependant posée : les données doivent être anonymisées de sorte que les patients ne puissent être identifiés – l'OFS procède déjà ainsi de manière systématique pour sa statistique médicale.

Soucieux de rendre disponibles de meilleures bases décisionnelles pour l'assurance-maladie et la santé publique, nous avons préparé en tenant compte de nos travaux précédents, avec la collaboration de l'Institut d'économie et de management de la santé (IEMS) de l'Université de Lausanne, les bases méthodologiques et les données techniques permettant de procéder à des analyses de données détaillées. Dans un deuxième temps, nous avons validé à l'aide de trois pathologies importantes (tumeurs, diabète et maladies psychiques) les méthodes permettant d'exploiter les données de masse de manière appropriée et efficace.

Grâce au soutien du service de cryptologie du Département fédéral de la défense, de la protection de la population et des sports (DDPS), les problèmes complexes en matière de protection des données ont pu être analysés en profondeur et résolus, puisque les fichiers de données concernant les patients peuvent être associés à un code de liaison qui garantit l'anonymat complet des assurés.

La présente analyse montre que le développement des indicateurs basés sur le diagnostic a été réalisé avec relativement peu de moyens. Une extension ciblée des bases décisionnelles en matière de santé publique est ainsi étayée et celles-ci peuvent être utilisées à des fins diverses. Elles servent, par exemple, à estimer l'efficacité, l'adéquation et le caractère économique des prestations de l'assurance-maladie, à mieux cerner et chiffrer les problèmes principaux de la santé publique ainsi que le besoin et l'efficacité de mesures préventives et de traitements médicaux. Elles peuvent également servir de soutien aux cercles de qualité en médecine, orienter plus concrètement la formation et, par la même occasion, être utilisées pour renforcer le caractère économique et scientifique des traitements.

## Vorwort

Seit Jahren gilt das schweizerische Gesundheitssystem als statistisch unerschlossen, als Entwicklungsland mit wenig Hoffnung auf baldige Besserung. Dabei wird geflissentlich übersehen, dass die verfügbaren statistischen Angaben wesentlich mehr Analysen zulassen als heute realisiert werden und sich die vorhandenen Verwaltungsdaten der Krankenversicherer mit geringem Aufwand für eine kostengünstige Verbesserung der heutigen, unbefriedigenden Situation nutzen liessen. Naheliegend ist, dass sich die im Rahmen der Rechnungsstellung automatisch anfallenden Daten nutzen liessen:

- die Angaben zur den versicherten Personen (Alter, Geschlecht etc)
- die Rechnungsangaben über die bezogenen Leistungen (u.a. Rechnungsbetrag, Kostenbeteiligung, Leistungsart, Leistungserbringer)
- die Angaben zu den Bezüglern von Medikamenten und die dazugehörigen Pharmacodes

Durch die Kombination dieser Angaben mit den Angaben der medizinischen Statistik über die Diagnosen und Interventionen der Behandlungen im stationären Bereich lassen sich Behandlungspfade abbilden und so zusätzliche Informationen ableiten. Voraussetzung ist allerdings, dass die Daten so anonymisiert werden, dass eine Identifikation der Patienten ausgeschlossen ist, so wie dies bereits heute systematisch bei der Medizinischen Statistik des BFS gemacht wird.

Im Bestreben bessere Entscheidungsgrundlagen für die Krankenversicherung und das Gesundheitswesen bereitzustellen, erarbeitete das BAG - aufbauend auf früheren Arbeiten des Amtes - zusammen mit dem Institut de la santé (IEMS) der Universität Lausanne die methodischen und datentechnischen Grundlagen um detaillierte Datenanalysen vornehmen zu können. In einem zweiten Schritt wurde dann anhand von drei wichtigen Pathologien (Krebs, Diabetes und psychische Erkrankungen) mit einer Machbarkeitsstudie geprüft, ob mit den erarbeiteten Methoden Massendaten zweckmässig und effizient verarbeitet werden können.

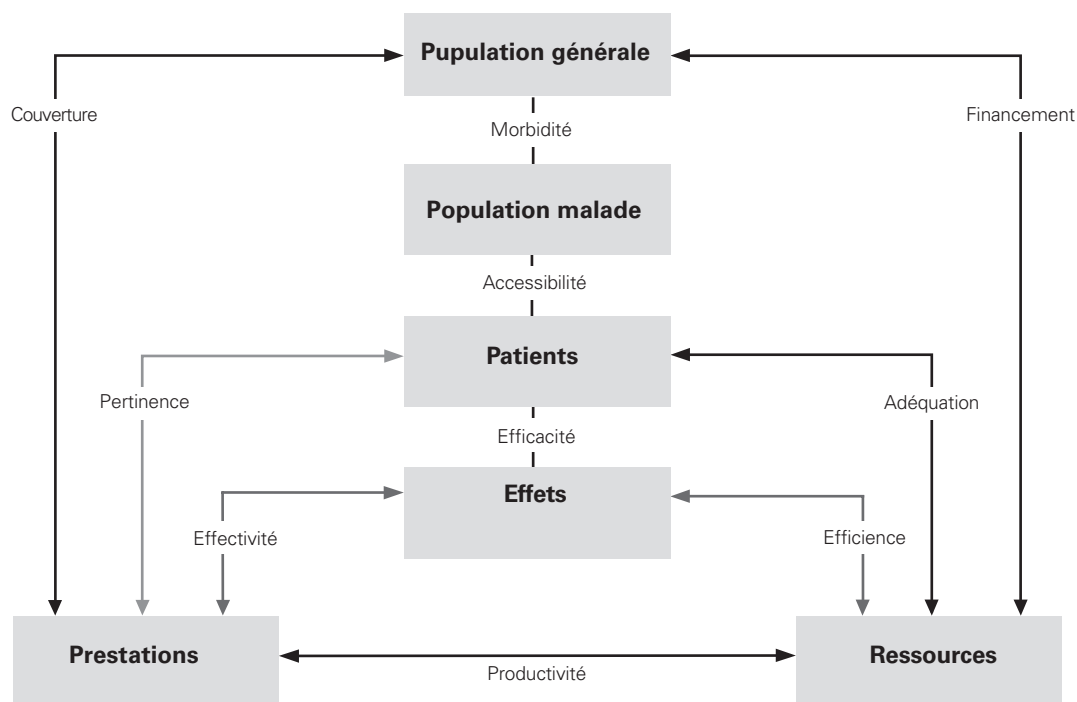
Dank der Unterstützung durch den Dienst für Kryptologie des VBS konnten die mit den vorgesehenen Analysen verbundenen, schwierigen Datenschutzprobleme eingehend analysiert und dann gelöst werden, indem die Datenbestände auf der Patientenebene mit einem Verbindungscode kombiniert werden, der die vollständige Anonymität der Versicherten gewährleistet.

Die vorliegende Analyse zeigt, dass diagnosegestützte Indikatoren mit verhältnismässig geringem Aufwand machbar sind. Damit ist die Grundlage gelegt für einen gezielten Ausbau der Entscheidungsgrundlagen des Gesundheitswesens, die für unterschiedliche Zwecke eingesetzt werden können. Neben der Beurteilung der Wirksamkeit, Zweckmässigkeit und Wirtschaftlichkeit der Krankenversicherungsleistungen könnten u.a. die Hauptprobleme der öffentlichen Gesundheit und damit der Bedarf und die Wirksamkeit präventiver Massnahmen und medizinischer Behandlungen besser erfasst und beziffert werden. Sie könnten aber auch zur Unterstützung medizinischer Qualitätszirkel und zur besseren Abstützung der medizinischen Ausbildung auf Evidenznachweis und Kosten-Nutzenüberlegungen eingesetzt werden.



## Résumé

Sur la base d'une analyse de la littérature et d'une revue systématique des indicateurs existants dans d'autres pays, une dizaine d'indicateurs ont été définis dans le but de suivre les performances du système de santé, notamment la couverture et le financement des prestations, la morbidité et l'accès aux soins, la justification des prestations, le coût des prises en charge et la mesure des effets obtenus. Ces indicateurs sont construits en mettant en relation les données relatives à six entités : population, malades, patients, prestations, ressources et effets :



Chacune de ces entités peut être quantifiée (nombre de personnes pour la population, francs pour les ressources, points pour les prestations par exemple) et qualifiée (âge et sexe de la population, nature des maladies, type de prestation, type d'effets – guérison, satisfaction, survie, effets indésirables –, qualification des professionnels ou nature des ressources par exemple).

Les indicateurs sont construits en mettant en relation deux entités, souvent sous la forme de ratios ou de taux. Les indicateurs de morbidité sont généralement définis par la densité d'incidence, qui exprime la vulnérabilité d'une population à une certaine maladie, ou par la prévalence qui indique la proportion d'une population dans un état de santé donné. La couverture met en relation des prestations avec une population cible, par exemple dans le cadre d'une campagne de vaccination ; elle est exprimée en quantité de prestations par personne visée.

Les indicateurs de financement mesurent la quantité de ressources consommées par personne ou selon leurs sources de financement.

L'accessibilité est la propension des malades à devenir des patients ; elle peut être mesurée spatialement (distances) ou temporellement (délais d'attente). Les indicateurs de pertinence mettent en relation les prestations et les caractéristiques des patients (case mix), alors que les indicateurs d'adéquation mettent en relation la quantité de ressources consommées par ces patients. La mesure des effets reflète les objectifs du système de santé, à savoir répondre aux attentes des patients (efficacité), vérifier que les prestations produisent vraiment les effets attendus (effectivité) et que les ressources sont allouées de manière à maximiser les effets (efficience).

### Les données sont toutes disponibles

La production des indicateurs retenus devrait pouvoir se faire à un coût modeste, parce qu'ils sont tous basés sur des données déjà disponibles (tableau 1).

#### *Données exigées*

Données sur les assurés	âge, sexe, canton, assureur, début et fin de validité du contrat d'assurance, frais hospitaliers et ambulatoires à charge de l'assurance, participation financière de l'assuré, montant de la franchise contractuelle, décès éventuel
Données de facturation	fournisseur de la prestation, prescripteur, date de facturation, quantité de contacts, type de prise en charge (hospitalier, etc.), montant, pharmacode et renouvellement de la délivrance (s'il s'agit de médicaments)
Statistique médicale des hôpitaux	diagnostics, opérations, transfert avant ou après le séjour, durée de séjour, poids à la naissance (s'il s'agit de nouveau-nés)

Les données administratives sur les assurés et les factures sont les mêmes que celles qui avaient été demandées lors de l'élaboration de la LAMal (1996). Les données sur les médicaments sont aujourd'hui déjà envoyées sous forme informatisée aux assureurs par les organismes de facturation des pharmacies, et les mêmes informations devraient être fournies par les médecins dispensateurs de médicaments par l'intermédiaire des «centres de confiance» agréés par la FMH. Cette manière de procéder permet d'éviter la mise sur pied d'un recueil coûteux de nouvelles données par exemple sur les diagnostics des patients ambulatoires. L'ordonnance sur l'assurance-maladie obligatoire précise par ailleurs aux articles 28, 30, 31 et 32 de l'ordonnance de la LAMal que les assureurs-maladie sont tenus de livrer leurs données à l'OFSP. La mise en oeuvre peut donc se faire sur la base de la législation en vigueur. Quant à la statistique médicale des hôpitaux, elle est limitée aux hospitalisations ; elle est obligatoire pour tous les hôpitaux et cliniques suisses et gérée depuis 1998 par l'Office fédéral de la statistique.

### Anonymat et protection des données

Pour garantir l'anonymat des patients les données doivent être cryptées et retraitées de manière à ce qu'une identification des assurés ne soit plus possible même par un couplage avec d'autres sources de données. Dans une deuxième partie du rapport les procédures nécessaires pour assurer ce but sont analysées par le Service de cryptologie de l'Armée suisse.

### Test de faisabilité

Un test de faisabilité a été effectué par l'IEMS avec un groupe limité d'assureurs pour un échantillon de 170 000 assurés de trois cantons (données 2002). Des indicateurs ont été construits pour trois pathologies courantes : le cancer, le diabète et les maladies mentales. Pour identifier ces pathologies, trois informations peuvent être utilisées : les diagnostics fournis par la statistique médicale des hôpitaux, la spécialité du médecin traitant (pour la psychiatrie) et la nature du traitement médicamenteux prescrit (prise régulière d'insuline pour le diabète par exemple). Le test de faisabilité a démontré plusieurs points :

- tous les indicateurs prévus ont pu être calculés;
- les données médicales hospitalières sont très utiles dans la mesure où la majorité des assurés coûteux passent à un moment ou à un autre à l'hôpital;
- il est souvent possible d'identifier les maladies sur la base des médicaments dispensés, sauf pour le cancer ; l'intégration ultérieure des registres dans le système d'information pourrait être envisagée;
- les données aujourd'hui disponibles concernant la dynamique des contacts avec les services de soins permettraient de développer des indicateurs nouveaux d'efficacité des soins (survie à moyen terme, pourcentage de guérison, etc.).

## Exemples d'indicateurs

Indicateurs		Canton I	Canton II	Canton III
couverture	Nombre de consultations psychiatriques/ nombre d'assurés	0.20	0.32	0.45
financement	Coût annuel d'un assuré selon son état de santé			
	– d'un cancer		12'900 Fr.	
	– d'un diabète		4'600 Fr.	
	– d'une affection mentale		6'400 Fr.	
accessibilité	pour-cent de patients cancéreux ayant bénéficié de services à domicile	14 %	19 %	28 %
	idem pour des patients en fin de vie	4.7 %	5.2 %	7.3 %
pertinence	pour-cent de patients psychiatriques consommant des somnifères			
	– femmes	15 %	23 %	31 %
	– hommes	11 %	22 %	24 %
pertinence	pour-cent de diabétiques			
	– ayant consulté un ophtalmologue dans l'année (recommandé)	41 %	34 %	36 %
	– ayant reçu un traitement par les thiazolidinediones (discuté)	8 %	16 %	13 %
productivité	Coût par consultation psychiatrique	149 Fr.	109 Fr.	157 Fr.
adéquation	Coût annuel d'une leucémie ou d'un lymphome	14'000 Fr.	12'300 Fr.	21'800 Fr.
	– dont coût hospitalier	10'140 Fr.	6'310 Fr.	11'822 Fr.
	– médicaments	1'780 Fr.	3'222 Fr.	4'233 Fr.
effectivité	Proportion de patients diabétiques hospitalisés pour un diabète compliqué	0.020	0.052	0.019
efficacité	Proportion de cancers du sein opérés de manière conservatrice	67 %	35 %	38 %

Sur les 170'000 assurés suivis, les environs 100 000 qui sont tombés malades ont généré plus de 700'000 factures et se sont vu prescrire plus d'un million de médicaments.

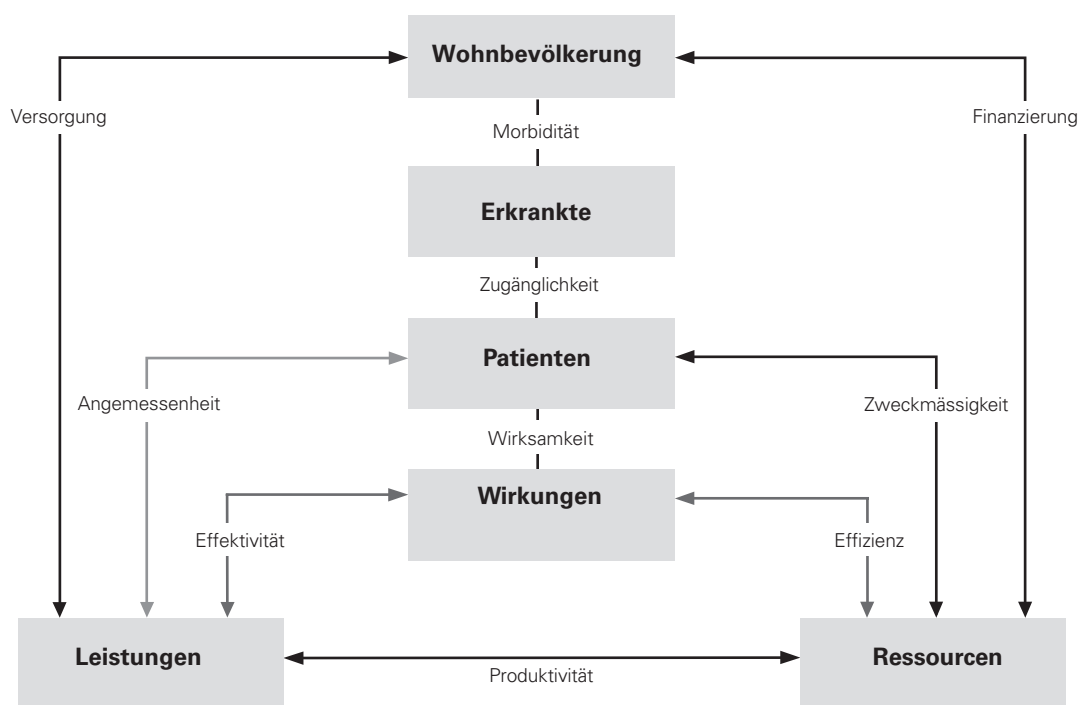
Plusieurs indicateurs ont été calculés pour illustrer la démarche. Les résultats chiffrés doivent être interprétés avec prudence, parce que l'échantillon des patients n'est pas représentatif et que le niveau de signification statistique n'a pas été calculé. Il est donc possible que les valeurs soient biaisées ou dues au hasard. Ils sont cependant fournis ici à titre d'illustration pour en montrer l'intérêt.

L'étude de faisabilité montre cependant de fortes variabilités régionales dans la couverture des besoins, la prescription des médicaments et les coûts des prestations médicales. Leur analyse est actuellement difficile car aujourd'hui on dispose seulement des dépenses moyennes par assuré. Avec les indicateurs proposés il sera dorénavant possible de répartir les coûts selon les différentes maladies et leur gravité, ce qui permettra d'analyser l'économicité des traitements sur la base des coûts par maladie. En même temps il sera enfin possible de mieux montrer le succès et l'utilité des traitements et non seulement les coûts générés. Cela permettra de mettre en évidence des différences de pratique et d'alimenter une discussion constructive entre médecins praticiens et médecins conseils des caisses-maladie.

Enfin, une innovation principale réside dans la possibilité de mesurer l'impact des soins. On pourrait ainsi vérifier que les soins ambulatoires sont de qualité et ont été fournis à temps, mesurant l'incidence des hospitalisations potentiellement évitables (comas diabétiques, asthme, certaines infections extensives par exemple). Les analyses à plus long terme permettraient également de vérifier le pronostic de ces patientes (survie à cinq ans). L'étude de faisabilité laisse augurer de nouveaux indicateurs fondés sur le devenir des patients (proportion de patients guéris, soulagés ou qui ont vu au contraire leur maladie s'aggraver).

## Zusammenfassung

Gestützt auf eine Literaturanalyse und eine systematische Durchsicht der in anderen Ländern verwendeten Indikatoren werden zehn Indikatoren vorgeschlagen, um die Leistungen des Schweizerischen Gesundheitssystems, insbesondere die Deckung und die Finanzierung der Leistungen, die Morbidität und den Zugang zur Gesundheitsversorgung, die Rechtfertigung der Leistungen, die Behandlungskosten und die Messung der erzielten Wirkungen, zu erfassen. Die Indikatoren werden gebildet, indem sechs Grössen (Bevölkerung, Erkrankte, Patientinnen und Patienten, Leistungen, Wirkungen und eingesetzte Mittel) zueinander in Bezug gesetzt werden:



Jede dieser Grössen kann quantitativ (z.B. Anzahl Personen für die Bevölkerung, Franken für die eingesetzten Mittel, Tarifpunkte für die Leistungen) und qualitativ (z.B. Alter und Geschlecht der Personen, Art der Erkrankungen und Leistungen, Art der Wirkungen – Genesung, PatientInnenzufriedenheit, Lebenserwartung, Nebenwirkungen – Qualifikation der Berufspersonen oder Art der eingesetzten Mittel) bewertet werden.

Für die Bildung von Indikatoren werden je zwei Grössen in Form einer Verhältniszahl oder einer Quote zueinander in Beziehung gesetzt. Die Morbiditätsindikatoren basieren in der Regel auf der Inzidenz, welche das Auftreten einer Erkrankung in einer Bevölkerungsgruppe ausdrückt, oder auf der Prävalenz, welche den Anteil einer Bevölkerungsgruppe mit einem bestimmten Gesundheitszustand angibt. Für die Messung der Versorgung werden die Leistungen einer Zielgruppe gegenübergestellt, beispielsweise im Rahmen einer Impfkampagne. Ausgedrückt wird der Erfassungsgrad mit dem Verhältnis zwischen der Anzahl Leistungen und der Anzahl Zielpersonen.

Die Finanzierungsindikatoren ihrerseits messen die Menge der eingesetzten Mittel pro Person bzw. nach Finanzierungsquelle.

Die Zugänglichkeit ist eine Messgrösse, um die Wahrscheinlichkeit, dass eine erkrankte Person ambulant oder stationär behandelt wird, zu erfassen. Der Zugang zur Gesundheitsversorgung kann räumlich (Distanzen) oder zeitlich (Wartezeit) gemessen werden. Die Indikatoren zur Zweckmässigkeit setzen die Leistungen mit den Merkmalen der behandelten Personen in Bezug (Casemix), wogegen die Indikatoren zur Messung der Angemessenheit das Verhältnis der verbrauchten Mittel pro behandelte Person

messen. In der Analyse der Wirkungen widerspiegeln sich die Ziele des Gesundheitssystems: Werden die Erwartungen der Patientinnen und Patienten erfüllt (Wirksamkeit), erzeugen die Leistungen die erwarteten Wirkungen (Effektivität), werden die Mittel so eingesetzt, dass die grösstmögliche Wirkung erzielt wird (Effizienz)?

### **Die Datengrundlagen sind bereits vorhanden**

Die Erstellung der gewählten Indikatoren verursacht keine grossen zusätzlichen Kosten, da alle Indikatoren auf bereits vorhandenen Daten basieren.

#### *Benötigte Daten*

Daten über die Versicherten	Alter, Geschlecht, Kanton, Versicherer, Beginn und Ende der Gültigkeit des Versicherungsvertrags, Kosten der stationären und ambulanten Behandlung zulasten der Versicherung, finanzielle Beteiligung der versicherten Person, Franchisehöhe, allfälliger Todesfall.
Rechnungsdaten	Leistungserbringer, Rezeptaussteller, Rechnung, Anzahl Konsultationen, Art der Behandlung (stationär usw.), Betrag, Pharmacode und abgegebene Menge (falls es sich um Medikamente handelt).
Medizinische Statistik der Krankenhäuser	Diagnosen, Operationen, Verlegung vor oder nach dem Aufenthalt, Aufenthaltsdauer, Geburtsgewicht (bei Neugeborenen). Die Verknüpfung dieser Daten wird durch einen anonymen Verbindungscode gewährleistet (vgl. Erläuterungen im Text).

Die Daten über die Versicherten und die Rechnungen für bezogene Leistungen sind bei den Krankenversicherern vorhanden. Die Daten über die Medikamente werden bereits heute von den Apotheken in elektronischer Form an die Versicherer gesandt. Die selbstdispensierenden Ärztinnen und Ärzte liefern die entsprechenden Informationen meist über die von der FMH anerkannten «TrustCenter» auch in elektronischer Form. Mit der vorgeschlagenen Lösung lässt sich daher vermeiden, dass kostspielige neue Datenerhebungen, etwa über die Diagnosen von ambulant behandelten Personen bei den Leistungserbringern eingefordert werden müssen. In den Artikeln 28ff, der Verordnung über die obligatorische Krankenversicherung, ist im Übrigen vorgesehen, dass die Krankenversicherer dem BAG ihre Angaben zu übermitteln haben. Die Umsetzung kann somit auf der Grundlage der geltenden Gesetzgebung erfolgen. Die Medizinische Statistik der Krankenhäuser ist auf die Spitalaufenthalte beschränkt; sie ist für alle Krankenhäuser und Kliniken in der Schweiz obligatorisch und wird seit 1998 vom Bundesamt für Statistik geführt.

### **Anonymität und Datenschutz**

Um die Anonymität der behandelten Personen zu gewährleisten müssen die Daten verschlüsselt und so aufbereitet werden, dass eine Identifizierung der einzelnen Versicherten auch wenn sie mit anderen Daten zusammengeführt würden, unmöglich ist. Das dazu notwendige Vorgehen und die Fragen der Verschlüsselung und des Hashings wurden durch den Dienst für Kryptologie der Schweizer Armee detailliert analysiert.

### **Machbarkeitsstudie**

Anhand der Daten aus dem Rechnungsjahr 2002 hat das IEMS bei einer Gruppe von Versicherern mit einer Auswahl von rund 170 000 Versicherten aus drei Kantonen eine Machbarkeitsstudie durchgeführt. Für drei verbreitete Krankheitsbilder – Krebs, Diabetes und psychische Erkrankungen – wurden die im ersten Teil der Forschungsarbeit vorgeschlagenen Indikatoren gebildet. Zur Bestimmung und Zuordnung der Krankheiten können drei Angaben verwendet werden: die Diagnosen aus der Medizinischen Statistik der Krankenhäuser, der Facharztstitel der/des behandelnden Ärztin/Arztes und die Art der verschriebenen medikamentösen Behandlung (z.B. regelmässige Einnahme von Insulin bei Diabetes).

**Die Machbarkeitsstudie hat folgende Ergebnisse erbracht:**

- Alle vorgesehenen Indikatoren konnten berechnet werden.
- Die medizinischen Daten der Krankenhäuser sind sehr nützlich, da die meisten teuren Versicherten irgendwann einen Spitalaufenthalt verzeichnen.
- Die Krankheiten können häufig anhand der verabreichten Arzneimittel identifiziert werden, ausser bei Krebserkrankungen; eine spätere Einbindung der Krebsregister in das Informationssystem wäre daher nützlich und möglich.
- Mit Hilfe der heute verfügbaren Daten aus der Pflege könnten neue Indikatoren für die Wirksamkeit der Pflege gebildet werden (mittelfristige Überlebensrate, Anteil von Genesungen usw.).

*Beispiele für Indikatoren*

<b>Indikator</b>		<b>Kanton I</b>	<b>Kanton II</b>	<b>Kanton III</b>
Versorgung	Anzahl psychiatrischer Konsultationen/ Anzahl Versicherte	0.20	0.32	0.45
Finanzierung	Jährliche Kosten einer versicherten Person entsprechend ihres Gesundheitszustands			
	– bei einer Krebserkrankung		12'900 Fr.	
	– bei Diabetes		4'600 Fr.	
	– bei einer psychischen Erkrankung		6'400 Fr.	
Zugänglichkeit	% Krebspatientinnen/-patienten, die Pflege zu Hause erhalten haben	14 %	19 %	28 %
	idem für die Patientinnen/Patienten am Lebensende	4.7 %	5.2 %	7.3 %
Angemessenheit	% psychiatrischer Patientinnen/Patienten, die Schlafmittel einnehmen			
	– Frauen	15 %	23 %	31 %
	– Männer	11 %	22 %	24 %
Angemessenheit	% Diabetiker			
	– die im Untersuchungsjahr einen Augen- arzt konsultiert haben (empfohlen)	41 %	34 %	36 %
	– die eine Behandlung mit Thiazolidin- edionen erhalten haben (kontrovers)	8 %	16 %	13 %
Produktivität	Kosten je psychiatrische Konsultation	149 Fr.	109 Fr.	157 Fr.
Zweckmässigkeit	Jährliche Kosten einer Leukämie oder eines Lymphoms	14'000 Fr.	12'300 Fr.	21'800 Fr.
	– davon Spitalkosten:	10'140 Fr.	6'310 Fr.	11'822 Fr.
	– davon Medikamente	1'780 Fr.	3'222 Fr.	4'233 Fr.
Effektivität	Anteil Diabetespatientinnen/-patienten, die wegen einer komplexen Diabetes ins Spital eingewiesen wurden	0.020	0.052	0.019
Wirksamkeit	Anteil brusterhaltender Operationen bei Brustkrebs	67 %	35 %	38 %

Im Rahmen der vorliegenden Untersuchung wurden die Daten von 170'000 Versicherten aus drei Kantonen untersucht. Für die rund 100'000 während der Beobachtungsperiode erkrankten Personen wurden über 700'000 Rechnungen ausgestellt und über eine Million Medikamente verschrieben. Um die vorgeschlagene Vorgehensweise zu illustrieren und den Nachweis zu erbringen, dass es möglich ist die Indikatoren auch mit grossen Datenmengen zu berechnen, wurden die Indikatoren mit diesen Daten berechnet. Die zahlenmässigen Ergebnisse sind allerdings mit Vorsicht zu interpretieren, da die Patientenauswahl nicht mit einer repräsentativen Stichprobe gebildet und das statistische Signifikanzniveau

nicht ermittelt wurde. Die Resultate könnten daher verzerrt oder zufallsbedingt sein. Sie sind nur als Illustration aufgeführt, um ihren grundsätzlichen Nutzen aufzuzeigen.

Immerhin zeigt die Machbarkeitsstudie bereits deutlich, dass in den Bereichen Versorgung, Medikamentenverschreibung und bei der ärztlichen Versorgung erhebliche regionale Unterschiede bestehen. Deren nähere Untersuchung ist allerdings heute schwierig, denn zurzeit können die ärztlichen Leistungen nur anhand der Durchschnittskosten pro Konsultation ausgewiesen werden. Mit den vorgeschlagenen Indikatoren liessen sich die Merkmale der behandelten Personen und der Schweregrad der Erkrankungen mitberücksichtigen und so die Wirtschaftlichkeit medizinischer Behandlungen anhand der Kosten pro Erkrankung aufschlüsseln. So könnten auch die Behandlungserfolge und nicht nur die Kosten ausgewiesen werden. Damit liessen sich Unterschiede in den Behandlungspraktiken aufzeigen und Erkenntnisse für eine konstruktive Diskussion zwischen praktizierenden Ärzten und den Vertrauensärzten der Krankenkassen gewinnen.

Eine wichtige Verbesserung ergäbe schliesslich die Messung der Pflegewirkungen mit der sich überprüfen liesse, ob die ambulanten Pflegeleistungen kompetent und rechtzeitig erbracht wurden. Erfassbar wäre auch die Inzidenz potenziell vermeidbarer Spitalaufenthalte (z.B. diabetisches Koma, Asthma, gewisse ausgedehnte Infektionen). Untersuchungen über längere Zeiträume würden es zudem erlauben, Verlaufsprognosen einzelner Krankheiten (z.B. Fünfjahresüberlebensraten) zu überprüfen. Die Machbarkeitsstudie enthält schliesslich Anregungen für neue Indikatoren zu den Erkrankungsverläufen (Anteil Patienten, die geheilt wurden, deren Zustand sich gebessert hat oder aber bei denen sich die Krankheit verschlimmert hat).



# 1 Cadre conceptuel

## 1.1 Introduction

Les coûts de la santé s'élèvent en 2006 à 56 milliards de francs, 22 milliards sont imputables aux prestations de l'assurance-maladie obligatoire dont 15 milliards aux soins ambulatoires, 5 milliards en séjours hospitaliers et 2 milliards en établissements médico-sociaux. L'information disponible aujourd'hui est lacunaire pour suivre les performances du système de la santé, notamment dans le domaine ambulatoire; il n'est par exemple pas possible aujourd'hui de connaître pour chaque maladie le caractère économique et l'efficacité de la prise en charge des prestations de l'assurance-maladie obligatoire sur la base des données disponibles de routine. Une plus grande transparence permettrait à l'OFSP de mieux cibler sa politique de santé, aux assureurs d'établir des standards de comparaison sur une base objective et neutre, tandis que les fournisseurs de prestations (médecins, pharmaciens, services à domicile, physiothérapeutes, etc.) disposeraient d'informations précieuses pour améliorer leurs prises en charge.

Ce manque de vision globale sur l'utilisation des services de santé et les coûts par maladie n'est pas propre à la Suisse. Il peut être constaté dans la plupart des pays développés qui disposent pourtant souvent de données nombreuses sur les soins ambulatoires. Certes, de nombreuses études ont montré les variations régionales en information concernant les interventions chirurgicales, les dotations en médecins, l'usage de certains médicaments ou les techniques de diagnostiques. La plupart des acteurs du système de santé considèrent la réduction de telles variations en information comme importantes car elles ont un impact sur les coûts. Toutefois, il est difficile de conclure quant à leurs causes.

Des bases de données publiques ont également été mises sur pied : registres des causes de décès, registres des cancers ou des maladies contagieuses, statistiques médicales des hôpitaux, statistiques des établissements médico-sociaux et enquêtes sur l'état de santé de la population. De telles données sont couramment utilisées pour calculer des espérances de vie avec ou sans handicap, surveiller les risques épidémiologiques et comparer les systèmes de santé entre pays (indicateurs de l'OMS notamment). De tels indicateurs synthétiques sont intéressants et utiles pour mesurer les performances globales des systèmes de santé. L'évaluation de la performance d'un système de santé doit toutefois permettre d'attribuer la responsabilité d'effets observés à des secteurs spécifiques (au moins de façon conceptuelle). Des expériences spécifiques au domaine des soins ambulatoires ont été conduites dans certains pays, comme aux Etats-Unis, en se fondant sur des données médicales des dossiers des patients ou sur des enquêtes de satisfaction. Or, de telles données sont difficiles à collecter de routine et elles sont souvent ciblées sur quelques maladies seulement. En Suisse, les caisses-maladie ne disposent pas de données sur le profil médical des patients, ce qui permettrait de juger de la pertinence des prestations fournies. Or, la LAMal fixe des exigences précises quant aux remboursements des prestations, qui doivent être : économiques (peu coûteuses), adéquates (justifiées par le profil des patients) et efficaces (produisant les effets attendus). L'Organisation pour la coopération économique et le développement (OCDE) préconise par ailleurs de surveiller la capacité à répondre aux besoins des patients en terme d'amélioration de l'état de santé, ainsi qu'en terme d'accès équitable aux ressources. Elle demande également d'établir pour chaque indicateur des valeurs standard (benchmark).

Deux contraintes ont été ajoutées à ces objectifs, pour en faciliter l'acceptation en Suisse : les performances du système de santé devrait être aussi peu coûteuse que possible, c'est-à-dire basée sur des données disponibles de routine. En outre, la production des indicateurs doit garantir la protection de la sphère privée et le secret médical.

Le présent projet vise à définir les besoins en information de la Confédération, en mettant l'accent sur ceux qui ne sont pas satisfaits principalement dans le domaine des prestations pris en charge par LAMal et en particulier dans le domaine ambulatoire. Il va cependant de soi que le système d'information présenté ici sera également utile aux caisses-maladie pour leurs propres indicateurs et définir des valeurs standard pour la Suisse. Il est en effet généralement admis que le contrôle de gestion des médecins installés en cabinet est aujourd'hui lacunaire. Les médecins se plaignent par exemple de ne pas être



jugés en fonction de la lourdeur des affections présentées par leurs patients ainsi que du fait que leurs prestations en terme de santé ne sont pas considérées du point de vue sanitaire, mais seulement économique. Or, il faut bien se rendre compte que la tâche n'est pas aisée : les patients présentent souvent simultanément plusieurs maladies, ils sont suivis par plusieurs médecins dont les points de vue peuvent ou non se compléter mais qui génèrent souvent des prestations chez de multiples fournisseurs de soins. Il n'est donc pas possible, comme dans le domaine hospitalier par exemple, de définir clairement des produits avec un début (l'admission), une fin (la sortie) et un responsable de l'ensemble (l'hôpital).

En intégrant les données provenant de différentes sources et en reconstituant des épisodes de soins, il sera plus facile d'avoir une idée claire des conditions de prise en charge des prestations, selon les trois principes énoncés par la LAMal (article 32) à savoir : l'adéquation, le caractère économique et l'efficacité. Il sera ainsi possible de juger de la pertinence des prestations (principe de l'adéquation). Le calcul des coûts en sera également facilité, car ce sont l'ensemble des ressources consommées pour un même problème qui pourront être regroupées en un produit (principe du caractère économique). Enfin, le devenir du patient devrait pouvoir être étayé, pour s'assurer de l'efficacité des soins (principe de l'efficacité).

On peut espérer de ce système d'information des avantages pour toutes les parties concernées. Les caisses-maladie devraient pouvoir progressivement se doter d'instruments de contrôle plus adaptés, permettant une comparaison des pratiques médicales répondant aux critères suivants :

- décomposition des coûts par épisode de maladie ;
- calcul des valeurs attendues en fonction des problèmes médicaux posés par le patient ;
- analyse du devenir des patients ;
- analyse des profils de pratique (y compris sur la prévention).

Il est probable que le dispositif projeté ne réponde pas à toutes les attentes. Mais il s'agit d'un premier pas important, avec des investissements minimes par rapport aux enjeux économiques. Un exemple devrait illustrer l'esprit dans lequel se projet a été conçu. Il a souvent été question de demander aux médecins de fournir systématiquement les diagnostics de leurs patients ; or une telle démarche est relativement lourde et serait difficile à imposer au corps médical. Il paraît dès lors plus sage de se fonder sur les données déjà existantes (les statistiques médicales hospitalières et les prescriptions de médicaments), en les exploitant au mieux. Si après quelques années d'expérience, la preuve est faite que d'autres données devraient être récoltées (maladies ou traitements spécifiques), le système d'information pourra être étendu. On peut par exemple d'ores et déjà penser que les données des registres du cancer, les chimio- et radiothérapies, certaines prestations (positions TarMed spécifiques) pourraient être également intégrées ultérieurement au système d'information.

Pour les médecins, le fait d'être jugés sur un système qui reflète mieux la dimension médicale de leurs pratiques devrait également présenter des avantages. On peut notamment penser à :

- l'appui d'indicateurs objectifs pour situer leur propre pratique par rapport à celle de leurs collègues, permettant par exemple de mieux cibler les besoins en formation continue ;
- la baisse des demandes inutiles de renseignements par les assureurs, l'analyse tenant dorénavant mieux compte des maladies présentées par leurs patients (analyses moins biaisées) ;
- un travail facilité pour les médecins conseils qui disposeraient de filtres statistiques permettant d'isoler les pratiques les plus atypiques (disparition des demandes «au hasard»).
- une plateforme d'informations utiles pour les discussions au sein des cercles de qualité.

Une meilleure information partagée entre les caisses-maladie et les médecins devrait également permettre d'affiner les instruments de gestion de la LAMal. Ainsi, le débat sur la suppression de l'obligation de contracter – qui est une mesure du type «tout ou rien» - pourrait évoluer vers des solutions plus nuancées, par exemple en instaurant des valeurs du point TarMed différenciées selon que le médecin respecte les critères définis par la LAMal (l'adéquation, le caractère économique et l'efficacité des pres-

tations). De même, la clause du besoin – qui limite le nombre de médecins installés en cabinet privés pourrait être remplacée par une analyse du marché médical par spécialité, qui tienne compte du volume de consultations et pas seulement de l'enregistrement d'un médecin dans le registre des créanciers (problème des temps partiels, fin d'activité en pratique privée, etc.). L'instauration d'une information est clairement à l'avantage des caisses-maladie, car elle devrait permettre d'éviter des gaspillages. Mais elle sert également les fournisseurs de prestation, en évitant qu'une partie d'entre eux accaparent une trop forte quantité de ressources en multipliant des prestations inutiles. L'expérience récente de l'enveloppe budgétaire ambulatoire (la «neutralité TarMed») a montré qu'une limitation des prestations facturées permet de maintenir un prix du point plus élevé.

Il est important de noter ici que le système projeté ne se substitue pas à celui qui existe entre les fournisseurs de prestations et les caisses-maladie. Mais il instaure un dispositif d'information neutre et objectif, sur la base d'une partie seulement des données qu'ils échangent : essentiellement des données anonymisées sur les patients, les montants et les dates des factures, ainsi que les prescriptions médicamenteuses.

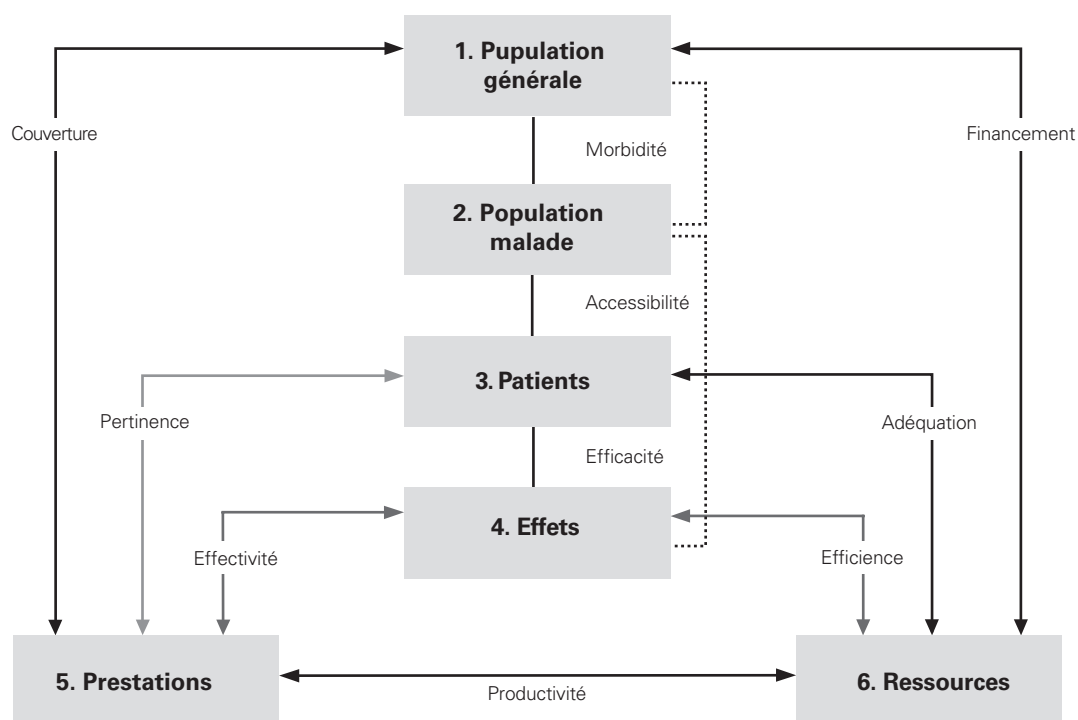
Sur la base d'une analyse de la littérature et d'une analyse systématique des indicateurs existants dans d'autres pays, une dizaine d'indicateurs ont été définis dans le but de suivre les performances du système de santé, notamment la couverture et le financement des prestations, la morbidité et l'accès aux soins, la justification des prestations, le coût des prises en charge et la mesure des effets obtenus. Pour plus de détails, le lecteur se reportera à l'article mentionné plus haut.

Ces indicateurs sont construits en mettant en relation les données relatives à la population, aux malades, aux patients, aux prestations, aux ressources et aux effets (figure 1). La population comprend toutes les personnes résidant en Suisse. Les malades sont les personnes atteintes par une maladie, qu'elles en soient conscientes ou non, que le diagnostic ait été posé ou non. Les patients sont les personnes ayant au moins un contact avec un fournisseur de soins. La distinction entre les prestations et les ressources permet de décrire les coûts de production. La notion d'effet a été préférée ici à celle de résultat, parce qu'il paraît essentiel d'établir une relation causale entre les composantes du système de soins et les résultats observés. Les effets peuvent être objectifs (décès) ou subjectifs (satisfaction des patients), attendus (amélioration de l'état de santé, soulagement de la douleur, etc.) ou inattendus (effets secondaires, complications, autres incidents). Chacune de ces entités peut être caractérisée par des descripteurs qualitatifs ou quantitatifs. Par exemple, les ressources peuvent être quantifiées en équivalents plein temps ou en terme monétaire, mais elles peuvent également être caractérisées par la qualification du personnel. Les indicateurs sont construits en mettant en relation plusieurs descripteurs, souvent sous la forme de ratio : par exemple, la productivité peut être mesurée par le nombre de points TarMed par franc consommé.

Les indicateurs de morbidité sont généralement définis par la densité d'incidence, qui exprime la vulnérabilité d'une population à une certaine maladie, ou par la prévalence qui indique la proportion d'une population dans un état de santé donné. La couverture met en relation des prestations avec une population cible, par exemple dans le cadre d'une campagne de vaccination ; elle est exprimée en quantité de prestations par personne visée. Les indicateurs de financement mesurent la quantité de ressources consommées par personne. L'accessibilité est la propension des malades à devenir des patients ; elle peut être mesurée spatialement (distances) ou temporellement (délais d'attente). Les indicateurs de pertinence mettent en relation les prestations et les caractéristiques des patients (case mix), alors que les indicateurs d'adéquation mettent en relation la quantité de ressources consommées par ces patients. La mesure des effets reflète les objectifs du système de santé, à savoir répondre aux attentes des patients (efficacité), vérifier que les prestations produisent vraiment les effets attendus (effectivité) et que les ressources sont allouées de manière à maximiser les effets (efficience).

Ce modèle repose entièrement sur des données déjà existantes, mais aujourd'hui éparses. Des techniques de cryptage et des mesures très strictes de protection des données ont été définies et mises en œuvre pour garantir l'impossibilité d'une identification d'un assuré ou d'un fournisseur de prestations.

Figure 1 Indicateurs



Les indicateurs liés à la population devraient permettre de mesurer les effets en terme d'espérance de vie en bonne santé, d'évaluer l'impact des campagnes de prévention en terme de morbidité évitée et d'autonomie sociale, de s'assurer de la couverture des besoins de la population telles que (visites médicales, de vaccination, de dépistage de cancers ou interventions discutables (lien avec les prestations) et de mesurer leur coûts (lien avec les ressources). Les indicateurs liés aux ressources intéresseront particulièrement les assureurs maladie, qui pourront ainsi s'assurer que les coûts sont couverts par le financement par les primes, les subsides fédéraux et cantonaux (lien avec la population), que les coûts sont conformes aux problèmes de santé de leurs assurés (lien avec les patients) et que les soins sont efficaces notamment. Quant aux fournisseurs de prestations, notamment les médecins, ils disposeront ainsi d'indicateurs pour justifier les prestations fournies (lien avec les patients), veiller à couvrir les besoins de la population et mesurer l'impact des soins fournis (lien avec les effets); les données recueillies devraient notamment servir à établir des valeurs standard, facilitant l'analyse de leurs pratiques dans le cadre des cercles de qualité et ainsi définir leurs éventuels besoins en formation continue. Les instances de défense des intérêts des patients devraient ainsi disposer d'informations sur l'accès aux soins, l'impact de ces soins, les risques encourus et peut-être influencer sur la politique de couverture des coûts (participation, franchise), soit tous des indicateurs liés aux patients. Enfin, on disposerait de données pour établir des liens entre les effets obtenus et les pratiques médicales et mesurer l'efficacité des soins. Comme on le voit, ces indicateurs fournissent une mesure objective permettant de réconcilier différents points de vue complémentaire dans notre système de santé. Ils couvrent par ailleurs les trois exigences fondamentales de la LAMal, qui charge la Confédération de contrôler le caractère économique des prestations fournies (indicateurs liés aux ressources), leur efficacité (indicateurs liés aux effets) et leur adéquation (indicateurs liés aux patients)<sup>1</sup>.

<sup>1</sup> On notera ici le double sens du terme «adéquation» qui désigne à la fois la justification des prestations par rapport au profil des patients (pertinence dans le modèle proposé ici) et le lien entre ces prestations et les ressources (dans le sens «adéquation» des moyens aux besoins des patients, tant en terme qualitatif (qualification du personnel par exemple) que quantitatif (montant en francs par exemple)).

Les données nécessaires au calcul des indicateurs sont pratiquement les mêmes que celles qui avaient été délivrées par les assureurs à l'OFSP pour la préparation de la dernière révision de la LAMal, auxquelles ont été ajoutées les données relatives aux prescriptions médicamenteuses. Ces dernières sont déjà envoyées sous forme informatisée aux assureurs par les organismes de facturation des pharmacies et les mêmes informations devraient être fournies par les médecins dispensateurs de médicaments par l'intermédiaire des «centres de confiance» agréés par la FMH. Cette manière de procéder permet d'éviter la mise sur pied d'un recueil coûteux de nouvelles données. L'ordonnance sur l'assurance-maladie obligatoire précise par ailleurs aux articles 28, 30, 31 et 32 de l'ordonnance de la LAMal, que les assureurs maladie sont tenus de livrer leurs données à l'OFSP. La liste des données demandées figure dans l'annexe A.

La majorité des indicateurs proposés sont nouveaux ; à cet égard, ils devront faire l'objet de validations scientifiques. La phase de développement impliquerait la mise en place de la base de données, la définition opérationnelle et la validation scientifique des indicateurs et différents contrôles relatifs à la protection des données qui a été effectués par le service de cryptologie de l'armée du DDPS avec la participation du représentant du préposé à la protection des données de la Confédération. Plus de détails sur ce sujet sont fournis dans l'article *Egglı Y., Halfon P., Chikhi M., Bandi T. Ambulatory healthcare information system: A conceptual framework. Health Policy 2006 ; 78 : 26-38*, qui fait partie intégrante du présent rapport.

### **Description du mandat**

L'une des principales lacunes des informations des prises en charge couvertes par la LAMal est que l'on ne dispose pas de données disponibles de routine sur les diagnostics. Les statistiques actuelles se limitent à la distribution des coûts par type de fournisseur, sans pouvoir les confronter au profil des patients touchés. Ces problèmes avaient notamment été identifiés lors d'études antérieures de l'OFSP (Evaluation des nouveaux modèles d'assurance-maladie) et de la Commission des statistiques sanitaires (CSS) notamment. Consciente de cette lacune, la Commission des statistiques sanitaires a décidé en 2003 de confier à un groupe de travail présidé par André Assimacopoulos des hôpitaux universitaires de Genève (HUG) l'initiative d'étudier la possibilité d'améliorer cette situation, en collaboration avec l'OFSP et ses partenaires.

Compte tenu de la multiplicité des sources de données et des contraintes liées à la protection des données, l'OFSP d'entente avec le groupe de travail et l'Institut d'économie et de management de la santé de l'Université de Lausanne (IEMS), ont eu comme mandat les objectifs suivants :

1. définir la liste des données nécessaires à la construction des indicateurs, notamment pour combler les lacunes du domaine des soins ambulatoires ;
2. clarifier les conditions d'obtention de ces données ;
3. éclaircir les questions relatives à l'anonymisation et la protection des données ;
4. proposer une démarche permettant de lier les prestations à des groupes de maladie et analyse des résultats.
5. tester la démarche proposée sur trois pathologies, à savoir les cancers, le diabète et les maladies mentales.

Pour réaliser ces objectifs, des données réelles provenant des assureurs maladie et de l'Office fédéral de la statistique ont été utilisées. Il a également été possible de valider les mesures prévues en matière de protection des données.

Les auteurs tiennent ici à remercier ici tous les membres du groupe de travail pour leurs critiques constructives et leurs suggestions :

André Assimacopoulos	Président du groupe de travail et de la CSS
Jean-Christophe Luthi	Directeur de l'Observatoire de la santé du valais
Judith Wagner	Hôpital cantonal d'Aarau.
François Weissbaum	Service de Cryptologie de l'armée du DDPS
Bettina Borisch	Professeur en Oncologie pathologie aux HUG de Genève
Jorge Garbino	Chef de clinique de Médecine interne aux HUG de Genève
Hans Peter Bickel	Spécialiste en assurance maladie
Clément Galopin	Spécialiste en assurance maladie
Pierre-Alain Tercier	Service de radio-oncologie de l'hôpital cantonal de Fribourg

Ils remercient également les organismes qui ont accepté de livrer leurs données, soit les assureurs maladie (remerciement anonyme), l'Office fédéral de la statistique (W. Weiss, A. Meister, A. Perret, P. Schwab), qui ont ainsi activement contribué à la réussite de ce projet.

Le présent rapport reprend les objectifs assignés par l'OFSP, dans l'ordre où ils viennent d'être énoncés. La seconde partie est consacrée aux exemples concrets, tirés des données récoltées. Enfin, ces travaux ont abouti à la publication d'un article scientifique paru dans la «Health Policy» : *Egglı Y., Halfon P., Chikhi M., Bandi T. Ambulatory healthcare information system: A conceptual framework. Health Policy 2006 (in press)*. Il fait partie intégrante du présent rapport et est résumé dans l'introduction du présent rapport ; pour plus détail et pour la revue systématique de la littérature scientifique, le lecteur s'y réfère.

## 1.2 Description des données

Les données proviennent d'un groupe d'assureurs-maladie ayant accepté de participer à la présente étude et de la Statistique médicale des hôpitaux. Un échantillon aléatoire de 169'992 assurés sous le régime de l'assurance-maladie obligatoire (LAMal), résidant en 2002 dans les cantons d'Argovie, du Tessin et de Vaud a été sélectionné. Parmi eux, 16'778 assurés ont été hospitalisés au moins une fois et 83'413 autres ont probablement souffert d'une maladie identifiée selon les critères décrits au chapitre 4. 715'654 factures de prestations ont été analysées pour ces 100'191 patients, alors que 1'180'047 médicaments dispensés<sup>2</sup>.

Les données fournies par les assureurs sont les suivantes :

Assurés :

- Numéro d'assuré crypté
- Age
- Sexe
- Canton de domicile
- Année d'assurance
- Début de validité (mois, année)
- Fin de validité (mois, année)

<sup>2</sup> En laissant de côté les médicaments de confort ou trop peu spécifiques pour suspecter une maladie.

## Cadre conceptuel

- Hospitalisation dans l'année (jour, mois, année)
- Séjour dans un EMS dans l'année (jour, mois, année)
- Décès éventuel durant l'année (oui/non)
- Coûts ambulatoires couverts par l'assureur
- Coûts séjours hospitaliers couverts par l'assureur

### Prestations :

- Numéro d'assuré crypté
- Numéro du fournisseur (numéro aléatoire, différent du numéro de créancier de santé suisse)
- Numéro du prescripteur (numéro aléatoire, différent du numéro de créancier de santé suisse)
- Date de la prestation (jour, mois, année)
- Montant

### Prescriptions médicamenteuses :

- Numéro de client crypté
- Numéro de prescripteur (numéro aléatoire, différent du numéro de créancier de santé suisse)
- Date de la délivrance (jour, mois, année)
- Pharmacode
- Quantité
- Montant

Le cryptage des numéros d'assuré a été effectué en utilisant la procédure mise en place par le service de cryptologie de l'armée du DDPS et de la section de la santé de l'Office fédéral de la statistique (OFS). Cette procédure permet d'attribuer un code anonyme permettant de lier ces données à celles d'autres sources de données telles que la statistique médicale des hôpitaux et les statistiques des instances de facturation des fournisseurs de prestations. Toutes les précautions ont en outre été prises pour qu'il ne soit pas possible d'identifier une personne sur la base de ses attributs. Les données relatives aux fournisseurs de prestations ont également été anonymisées, en suivant la procédure de protection des données décrites dans le chapitre 3.

Les données issues de la statistique médicale des hôpitaux sont les suivantes

- Numéro de patient crypté
- Numéro du prescripteur crypté
- Modes d'entrée et de sortie
- Durées de séjour (pour chaque séjour : jour, mois, année)
- Diagnostics (1 à 10), selon la Classification internationale des maladies, version 10
- Opérations (1 à 10), selon la Classification suisse des interventions chirurgicales (CHOP)
- Rang du séjour

Tableau 1 *Concordance entre la statistique médicale et les données des assureurs*

<b>Cantons</b>	<b>AG</b>	<b>TI</b>	<b>VD</b>	<b>Total</b>
Nombre d'assurés hospitalisés en 2002	7'288	4'969	4'521	16'778
dont statistique médicale de l'OFS	5'334	3'994	3'712	13'040
Exhaustivité	73 %	80 %	82 %	78 %

Ainsi, 78% des patients hospitalisés en 2002 d'après les numéros de créancier figurent également dans la statistique médicale des hôpitaux de l'OFS de cette même année<sup>3</sup>. Ce résultat montre qu'il manque environ 20 % des séjours hospitaliers. Quelles en sont les raisons ? Une partie de l'explication est certainement liée à un manque d'exhaustivité des relevés de la statistique médicale des hôpitaux en 2002. Il est cependant possible qu'une partie des non concordances soit due à la procédure de cryptage qui dépend de la qualité des données fournies (voir plus loin). Il est par ailleurs intéressant de noter que 2'879 hospitalisations de la statistique médicale des hôpitaux concernait des assurés sélectionnés pour lesquels aucune hospitalisation n'a été signalée par les assureurs. Il est probable qu'une partie de ces divergences soit liée à des problèmes de définition (hospitalisations de moins de 24 heures par exemple «hospitalisation de jour, etc.») ou que les frais y relatifs ne soient pas couverts par l'assurance-maladie obligatoire (assureurs LAA, assureurs privés, chirurgie esthétique ou autre participation directe des patients).

Ces imperfections ne sont pas gênantes dans le cadre de la présente analyse de faisabilité. Toutefois, elles nous incitent à tester une procédure de cryptage différente, fondée sur les numéros d'assurés plutôt que des dates de naissances. Ces dernières données ne seraient utilisées que pour suivre les personnes lorsqu'elles changent d'assureur ; ce dernier problème pourrait par ailleurs être résolu par l'introduction d'un numéro unique pour tous les assurés.

### 1.3 Anonymisation et protection des données

L'intégration des données aujourd'hui éparpillées dans plusieurs bases de données repose sur la possibilité de créer un code de liaison entre ces différentes bases qui garantisse l'anonymat et la protection des données.

Des efforts importants ont été nécessaires pour convaincre les partenaires concernés de participer à la présente étude, ce qui est rassurant sur le souci de chacun de protéger avec efficacité les données, qu'il s'agisse des patients ou des fournisseurs de prestations. L'expérience de la présente étude a par ailleurs permis de clarifier les procédures d'échanges de données qui avaient été sous-estimés lors de la rédaction des premiers protocoles.

Nous ne reviendrons pas sur les avantages du code de liaison anonyme, actuellement utilisé par la section de la santé de l'OFS dans le cadre des statistiques médicales et dont les principes sont décrits dans le présent rapport du service de cryptologie de l'armée du DDPS. Nous rappellerons simplement qu'il n'est pas possible de déduire l'identité d'une personne à partir de ce code de liaison anonyme, notamment en raison d'un processus de hachage.

Il est important de noter que si l'utilisation de cette procédure de hachage et de cryptage permet de garantir une anonymisation totale des personnes, la présence d'autres données sensibles ne permettent pas de retrouver un individu par recoupement. A l'évidence, ce risque a été évité ici puisque ces données sensibles ont été agrégées pour éviter justement ce type d'opération.

<sup>3</sup> Ce taux ne tient pas compte du fait qu'un même patient peut être hospitalisé plusieurs fois au cours d'une même année. Il désigne ici la proportion d'assurés hospitalisés au moins une fois en 2002 et pour lesquels au moins un séjour hospitalier a été codé dans la statistique médicale des hôpitaux.



Deux dangers subsistent toutefois :

- a) des recoupements plus subtils permettraient de cibler certaines personnes;
- b) les données de la base de données statistiques pourraient être fusionnées avec d'autres contenant des données sensibles.

Le premier risque est illustré par l'exemple suivant : un journaliste apprend qu'une personnalité connue est hospitalisée le 14 avril 2002 dans la clinique privée X. Connaissant son âge, la taille de l'hôpital et le jour de l'admission, le journaliste en question aurait de fortes chances d'identifier l'enregistrement correspondant dans la statistique médicale des hôpitaux. Théoriquement, le secret statistique peut être levé si l'on peut isoler deux enregistrements uniques dans une base de données; ce raisonnement s'appuie sur l'hypothèse que le chercheur pourrait être l'autre personne. Si l'on s'appuie sur ce raisonnement, il conviendrait de s'assurer que la base de données finalement disponible comporte toujours au moins trois enregistrements identiques.

Le second risque est illustré par l'exemple suivant : un assureur pourrait être tenté de fusionner ses propres données avec la base de données statistiques décrite dans le tableau de l'annexe A. Sur la simple base de l'enchaînement des dates de prestation, l'assureur serait par exemple en mesure d'identifier la majorité des individus de sa propre base de données et de lever ainsi l'anonymat.

Fort de ces constats, ainsi que des expériences conduites avec la collaboration de nos partenaires détenteurs de données, plusieurs exigences ont été impérativement respectées pour qu'une intégration ait pu être envisagée, tout en garantissant une protection efficace de ces données<sup>4</sup> :

1. La notion de secret statistique exposée plus haut ne devrait ainsi pas seulement s'appliquer à chaque fichier, mais être étendu à l'ensemble de la base de données mises en relation. Le risque mentionné plus haut disparaît du fait que leur utilisation doit remplir deux conditions : être préparées par le service de cryptologie de l'armée DDPS et être soumises au secret de fonction. Un nombre minimal de variables doit être recueillies de routine, pour limiter les risques de recoupement ; à court terme, ces données seraient les suivantes :
  - a) Assurés** : âge, sexe, canton de domicile, frais de séjours hospitaliers\*, frais ambulatoires\*, participation financière de l'assuré\*, franchise, présents en début/fin de période d'assurance ;
  - b) Prestations** : numéro anonyme de créancier, numéro anonyme du prescripteur, classe thérapeutique du médicament\*\*, type de prise en charge (hospitalisation, autre), date de la prestation\*\*\*, quantité de prestations, montant\*\*\* ;
  - c) Statistique médicale** : âge, sexe, canton de domicile, rang (ordre séquentiel) du séjour durant l'année, diagnostics\*\*, opérations\*\*, séjours avant/après l'admission, durée de séjour, poids à la naissance ;
  - d) Fournisseurs de soins** : numéro anonyme, type de fournisseur\*\* Cette liste de données pourrait être étendue ultérieurement, s'il est démontré qu'il n'en résulte pas de risques supplémentaires.
2. Les codes de liaison anonyme obtenus par hachage et cryptage des données doivent être renumérotés aléatoirement, pour effacer toute trace des opérations de cryptage ;
3. Les numéros anonymes des fournisseurs de prestations (créancier et prescripteur) sont attribués aléatoirement, de manière à effacer toute trace avec les numéros de créancier de santéuisse ;
4. Les variables suivies d'une étoile sont arrondies au millier, à la centaine ou à la dizaine

---

<sup>4</sup> Une validation serrée de chacun de ces points devra être effectuée ultérieurement au moment où la conception détaillée de cette statistique sera établie et que les rôles de chacune des parties concernées (OFSP, Service de cryptage de l'armée du DDPS, OFS, chargés d'accompagner la mise en œuvre de la statistique) auront été clarifiés.



5. Les variables suivies de deux étoiles sont agrégées en classes diagnostiques, opératoires et thérapeutiques (médicaments) ;
6. Les variables suivies de trois étoiles sont utilisées uniquement à des fins de calcul (constitution des épisodes de prestations, totalisation de montants partiels) et ne sont pas accessibles à des tiers (protection renforcée), sauf autorisation expresse selon les règles décrites plus bas (g);
7. Des croisements systématiques des variables disponibles (sans les variables à trois étoiles) sont effectués par assuré et par fournisseur de prestations, pour s'assurer qu'il existe toujours au moins trois enregistrements identiques ; si ce n'est pas le cas, les règles d'arrondi et de regroupements (classes) sont modifiées pour que cette condition puisse être garantie.

Pour une implémentation des procédures de relevées régulières de ces données, nous proposons qu'à ces procédures techniques s'ajoutent des mesures organisationnelles également cruciales pour assurer une protection efficace et des mesures comporteront au moins les exigences suivantes :

- a)** l'instance chargée de l'intégration des données doit être neutre, c'est-à-dire ne pas être liée à des enjeux liés à l'utilisation des données et ne pas être elle-même détentrice de données ; le service de cryptologie de l'armée du DDPS pourrait être cette instance neutre, au moins dans un premier temps ;
- b)** le cryptage et le hachage des données doivent être effectué par un instrument automatique directement par les fournisseurs des données ; les données ayant servi à cette opération ne doivent pas pouvoir être jointes à l'envoi ;
- c)** l'ensemble des données doivent être fournies à l'instance neutre citée plus haut, qui vérifie qu'aucune donnée non autorisée n'y figure ;
- d)** les techniques de hachage et de cryptage doivent avoir été préalablement mises à l'épreuve tant du point de vue de leur fiabilité que de leur sécurité ; elles sont livrées à l'instance neutre sans la documentation qui permettrait le cas échéant d'en dévoiler le mode opératoire ;
- e)** en tant qu'organe de contrôle de la LAMal, l'OFSP reçoit de l'instance neutre les données intégrées, sans les variables suivies de trois étoiles ; il peut mettre le cas échéant à disposition des données à d'autres utilisateurs, en suivant les règles usuelles (secret de fonction) ;
- f)** l'attribution des codes anonymes, les arrondis et les regroupements en classe sont appliqués de manière automatique pour éviter tout contact humain avec les données sensibles;
- g)** en principe, les données sont conservées (y compris celles suivies de trois étoiles) par l'instance neutre dans un lieu de haute sécurité, de manière à permettre la mise à jour de séries historiques; cette instance fixe également les règles exceptionnelles notamment pour des buts de recherche scientifique et peut le cas échéant faire expertiser la fiabilité et la sécurité du dispositif mis en place par le préposé à la protection des données.

Le code de liaison anonyme est aujourd'hui fondé sur une analyse phonologique (méthode «soundex») des noms et prénoms, ainsi que sur la date de naissance. L'opportunité de retenir plutôt un numéro d'assuré pour ce cryptage devrait être décidée en temps opportun en fonction de la disponibilité d'un tel numéro et de sa transmission pour l'analyse des prestations prises en charge par la LAMal. Il s'agira en effet de garantir la continuité avec la méthode de cryptage actuellement utilisée. Les tests effectués par cette étude ont montré qu'il est très important de valider la fiabilité du cryptage, autrement que par des contrôles de plausibilité très indirects. Le rapport fourni par le service de cryptologie de l'armée suisse du DDPS et faisant parti du présent rapport garanti tous les moyens de vérification de protection possible.

## 1.4 Méthode d'identification des maladies

Un test de faisabilité a été effectué sur la base des données d'un groupe d'assureurs et de la statistique médicale des hôpitaux en utilisant la procédure de cryptage du code de liaison anonyme mise au point par le service de cryptologie de l'armée du DDPS et actuellement utilisée par l'OFS. Seuls les assurés des cantons d'Argovie, du Tessin et de Vaud ont été pris en considération, pour l'année 2002. Les fournisseurs de prestations ont été anonymisés, après avoir déterminé leur nature (spécialité médicale, pharmacie, hôpital, etc., cf. exemples 1, 2, 3 en annexe).

Dans le but de promouvoir le modèle proposé, mais aussi d'identifier d'éventuels problèmes pratiques, l'OFSP a préparé en collaboration avec le groupe de travail et l'IEMS quelques exemples d'application ciblés sur trois pathologies courantes : le cancer, le diabète et les maladies mentales.

Pour identifier ces pathologies, trois informations peuvent être utilisées : les diagnostics fournis par la statistique médicale des hôpitaux, la spécialité du médecin traitant et la nature du traitement médicamenteux prescrit. Or ces critères ne sont pas toujours assez spécifiques pour attribuer un diagnostic à un patient. Par exemple, il ne suffit pas de consulter un oncologue pour être cancéreux. En revanche, la prise régulière d'insuline est un critère fiable pour identifier un diabète. C'est le taux d'hospitalisation pour l'affection dépistée qui a été retenu pour estimer la sensibilité de chaque critère de détection (tableau 2). La spécificité des traitements a été jugée a priori sur la base des indications fournies dans le Compendium suisse des médicaments.

Les différentes pathologies ont été identifiées sur la base des diagnostics fournis par la statistique médicale des hôpitaux, en recourant à la typologie SQLape®<sup>5</sup>, ainsi que des médicaments prescrits aux patients (cf. annexes B, C et D). Le choix de cette typologie est justifié par le fait qu'elle permet de refléter toutes les pathologies et interventions des patients, en attribuant plusieurs «étiquettes» en cas de besoin<sup>6</sup> (chez un quart des patients hospitalisés) Un autre avantage de cette typologie est qu'elle permet de hiérarchiser les problèmes en fonction de leur nature (sévérité, organe plus ou moins vital).

Les critères pour identifier les pathologies de l'étude ont été retenus au-delà d'un certain seuil du taux d'hospitalisation. Ainsi la détection des cas de cancers repose sur une hospitalisation motivée par un cancer, le recours à une chimiothérapie ou encore un traitement adjuvant (spécifique ou non)<sup>7</sup>. Par contre le critère consultation d'un oncologue ne paraît pas pertinent, parce que ces derniers prennent également en charge des patients relevant de la médecine interne non tumorale<sup>8</sup>. Cette approche sous estime vraisemblablement les cas : en effet les traitements chimiothérapeutiques délivrés directement par les hôpitaux sont ignorés. Les radiothérapies ne sont généralement pas documentées dans la statistique médicale des hôpitaux, mais ce problème pourrait être résolu par une participation volontaire mais systématique des radiothérapeutes à cette statistique<sup>9</sup>. L'utilisation des données des registres suisse des cancers permettrait en outre de valider cette approche et de calculer des indicateurs par types de cancer dans les cantons qui tiennent des registres.

---

5 [www.SQLape.com](http://www.SQLape.com). Les catégories ont été affectées sans prendre les opérations en considération.

6 Une typologie du type DRGs (Diagnosis Related Groups) n'attribue qu'une seule catégorie aux patients, en les subdivisant au besoin selon la présence ou non de co-morbidités ou de complications mais sans préciser la nature de ces dernières.

7 La prescription d'un anti-vomitif après une intervention chirurgicale n'est habituellement pas une indication ambulatoire (cf. annexe A).

8 Il n'est pas exclu que les consultations oncologiques soient prises ultérieurement en considération, lorsque des critères auront été proposés pour les autres maladies.

9 Une telle participation est actuellement à l'étude ; elle est pleinement appuyée par l'OFSP.

Tableau 2 Taux de patients hospitalisés pour la pathologie détectée

<b>Liste des pathologies détectées à priori</b>		<b>Taux</b>
Critères de détection du cancer		
1*	Chimiothérapie anticancéreuse des leucémies et lymphomes	28.0 %
2*	Chimiothérapie contre le cancer de la prostate	26.2 %
3*	Chimiothérapie contre le cancer du sein	18.9 %
4*	Chimiothérapie contre les autres cancers	16.7 %
5*	Traitement adjuvant spécifique	23.3 %
6*	Traitement adjuvant non spécifique (sauf 1 à 5)	60.0 %
7	Autre chimiothérapie non spécifique du cancer (sauf 1 à 6)	1.8 %
8	Consultation d'un oncologue (sauf 1 à 6)	9.0 %
9	Aucun des critères 1 à 6**	<1.0 %
Critères de détection du diabète		
10*	Traitement par de l'insuline	23.4 %
11*	Traitement par des hypoglycémiant oraux spécifiques du diabète	14.2 %
12	Traitement par des hypoglycémiant oraux non spécifiques du diabète (sauf 10 et 11)	0.0 %
13	Consultation d'un endocrinologue – diabétologue (sauf 10 et 11)	1.2 %
14	Aucun des critères 10 et 11**	<0.4 %
Critères de détection de maladies mentales		
15*	Traitement médicamenteux de troubles psychotiques	18.5 %
16*	Traitement médicamenteux de la dépression	10.5 %
17*	Consultation d'un psychiatre (sauf 15 et 16)	8.3 %
18	Traitement médicamenteux de l'anxiété (sauf 15 et 16)	4.6 %
19	Traitement par des somnifères (sauf 15 et 16)	5.3 %
20	Aucun des critères 15 à 17**	<2.2 %

\* critères retenus pour identifier les pathologies ci-dessus, dans la présente étude

\*\* il s'agit ici de taux maximaux, les assurés sains n'étant pas pris en considération ici.

La plupart des thérapies du diabète sont spécifiques. Le volume des ventes d'antidiabétiques a été largement utilisé pour évaluer la prévalence du diabète. Les taux d'hospitalisations plus élevés en cas d'insulinothérapie sont plausibles. Il serait également possible de déterminer avec une précision suffisante le type de diabète en utilisant l'âge au moment du diagnostic et le délai d'institution de l'insuline après le diagnostic, qui sont des informations déductibles des données de routine. Même si tous les patients diabétiques ne suivent pas un traitement médicamenteux, on peut considérer que la majorité des traitements du diabète peuvent être recensées pour être analysées.

Les maladies mentales sont plus délicates à isoler. D'une part, une seule prescription ne suffit pas à identifier un problème psychiatrique et les indications thérapeutiques ne sont pas toujours univoques. D'autre part, beaucoup de patients ne sont pas forcément traités avec des médicaments.

Pour la présente analyse, l'identification d'une maladie mentale requiert la présence d'au moins l'un des critères suivants :

- hospitalisation pour une affection mentale (catégorie SQLape) ;
- prescription d'un médicament antipsychotique ou antidépresseur au moins une fois chaque semestre de l'année ;
- au moins une consultation chez un psychiatre.

Cette manière de procéder évite en principe les cas «bagatelles» (consultation unique, prescription d'un généraliste non renouvelée, prise occasionnelle de somnifères ou d'anxiolytiques). Une typologie plus fine pourrait être adoptée ultérieurement, par exemple pour exclure les psychanalyses (rendez-vous réguliers à long terme).

Il est important de bien noter que les analyses présentées dans le présent rapport sont avant tout exploratoires et que des validations systématiques devront être effectuées pour s'assurer de l'exhaustivité des relevés, de l'absence de biais et de la précision des résultats (analyse statistique).

La détermination d'une maladie à partir de la médication est une opération délicate, qui doit être effectuée avec une grande rigueur. Quelques précisions sont ici indispensables concernant la méthodologie à suivre dans la perspective d'une généralisation de la démarche testée ici. Premièrement, les indications thérapeutiques sont celles qui ont été approuvées par l'office d'enregistrement des médicaments Swissmedic et officiellement reconnues sur le territoire suisse. Ceci permet de préciser le transcodage des «pharmacodes<sup>10</sup>» (produits autorisés à la vente en Suisse) dans la nomenclature internationale (codes ATC<sup>11</sup>). La connaissance du produit facturé permet en effet souvent de distinguer une ou plusieurs pathologies (cortisone par exemple), ainsi que les modes d'administration (per os ou pommade par exemple).

Cette nomenclature est préférable dans la perspective de comparaisons internationales, puisqu'elle a été mise en œuvre par l'Organisation mondiale de la santé (OMS). La classification est fondée sur une codification à 7 caractères, qui va jusqu'à la formule chimique. La correspondance entre les pharmacodes et les codes ATC nous a été fournie par les assureurs, mais une vérification soigneuse a été effectuée et a amené à quelques corrections de détail. Les transcodages ont été fondés sur les indications officielles des «Compendium des médicaments» des années 2001 à 2004. Elles ont été confrontées aux informations contenues dans le Codex Galenica/Document de 2000/01. Pour tenir compte de la variabilité des indications liées à la présentation du produit, nous avons codifié comme «non spécifique» l'ensemble des médicaments du groupe ATC (7) considéré, dans les cas où une forme galénique, un dosage, un mode d'administration peut avoir une indication différente de l'indication généralement reconnue. Dans certains cas nous avons pu tenir compte de l'information contenue dans le pharmacode et du sexe. Par exemple la triptoreline est utilisée chez l'homme dans l'indication cancer de la prostate mais ce même principe actif a des indications chez la femme dans la médecine de la reproduction.

La relation entre un médicament et une maladie n'est de loin pas univoque et l'inférence des indications thérapeutiques reconnues à une affection doit être réalisée avec beaucoup de prudence. D'une part l'indication thérapeutique peut n'apporter qu'une information très générale. Par exemple, un antibiotique peut être mis en relation avec une maladie infectieuse. Selon la molécule qu'il contient, il peut nous informer sur le type de germes en cause. Selon le dosage prescrit, il peut nous donner une idée sur le niveau de gravité de l'infection. Cependant un antibiotique à lui seul ne nous autorise généralement ni à nommer la maladie infectieuse en cause ni la partie de l'organisme concernée. Il y a toutefois quelques exceptions telle que le traitement unique des cystites aiguës par fosfomycine par exemple. D'autres part certaines

---

10 La Suisse a mis en place un mode de regroupement des produits thérapeutiques commercialisés (pharmacodes) qui lui est propre. Il prévoit 17 groupes thérapeutiques majeurs et il attribue un code à 6 chiffres à toutes les spécialités commercialisées. La systématique de ce regroupement tient compte soit de l'organe concerné soit de l'indication thérapeutique du principe actif. Son niveau de précision ne va cependant pas jusqu'au niveau du principe actif et un même code peut être attribué à plusieurs principes actifs différents. Nous trouvons ainsi dans des mêmes groupes des substances dont le mécanisme d'action et le profil thérapeutique sont très différents.

11 The Anatomical Therapeutic Chemical Classification System with Defined Daily Doses (ATC/DDD)

antibiothérapies ont pour indication une prophylaxie primaire : préopératoire, prévention de l'endocardite lors de procédures, prophylaxie d'infections opportunistes chez les immunodéprimés etc.

Une autre question qui devra être abordée avec soin est la relation entre affection et durée du traitement, dosage ou forme galénique des médicaments. Par exemple la détection d'une prescription d'un somnifère n'autorise pas le diagnostic d'un problème d'endormissement chronique. De même si le dosage délivré est très faible ou la prescription ponctuelle, nous pouvons envisager un objectif thérapeutique différent de celui de l'indication principale du produit : soulagement d'une douleur musculaire par une benzodiazépine dont l'indication principal est le traitement de l'anxiété. Nous trouvons sur le marché des produits qui offrent une ou plusieurs présentations, c'est-à-dire plusieurs dosages ou plusieurs formes galéniques. Une seule présentation pour un principe actif n'est pas nécessairement une preuve d'indication unique et spécifique (inhibiteurs TNF par exemple). Au contraire la présence de plusieurs formes galéniques ou de plusieurs dosages ne correspondent pas nécessairement à autant d'indications différentes. La prescription d'une forme galénique par rapport à une autre peut cependant orienter l'analyse sur une pathologie particulière. L'administration orale peut être destinée à une autre pathologie que celle reconnue pour la forme utilisable pour l'administration parentérale (méthotrexate). La fréquence des ordonnances, les doses prescrites sont également des informations intéressantes qui peuvent constituer autant d'indices sur la pathologie ou la gravité des symptômes, voire des phénomènes d'accoutumance. Nous n'en avons pas tenu compte dans la présente étude de faisabilité, par souci de simplicité. L'opportunité d'en tenir compte devra cependant être soigneusement pesée si des ambiguïtés devaient pouvoir être levées par ce moyen.

Un autre problème est lié à l'exhaustivité et à la précision des informations disponibles. Les prescriptions médicamenteuses ne sont pas systématiquement répertoriées de façon détaillée selon les différents canaux de prestations et de distribution. En effet, les officines publiques facturent nommément tous les médicaments ce qui permet de prendre en considération le nom de la spécialité, la forme galénique, le dosage et le nombre de doses prescrites. Il sera important de vérifier que les médecins dispensant directement des médicaments (en Suisse allemande principalement) fournissent également les mêmes données sur support informatique, selon les mêmes nomenclatures. Cette même règle devrait également s'appliquer aux hôpitaux et cliniques qui délivrent directement des médicaments (chimiothérapies par exemple). La question des médicaments dispensés en dehors des canaux de vente habituels (médications expérimentales par exemple) devra également être abordée.

Les cantons d'Argovie, du Tessin et de Vaud ont été choisis parce que les médicaments ne sont en principe dispensés que par les pharmacies. L'examen des factures relatives aux médicaments montre cependant qu'environ 10 % du chiffre d'affaires des médicaments n'est pas enregistré par des pharmacies dans ces cantons.

## 1.5 Avertissement méthodologique général

Idéalement, un indicateur devrait être pertinent, c'est-à-dire mesurer correctement le phénomène concerné. Tous les indicateurs proposés mettent en relation des caractéristiques d'entités distinctes, comme les prestations et les ressources par exemple. Deux conditions devraient à cet égard être observées :

- le domaine couvert par les deux entités devrait être strictement le même ;
- les mesures relatives à chacune de ces entités devraient être indépendantes.

Ces deux conditions ne sont pas toujours vérifiées en pratique. Par exemple, la morbidité du diabète n'est pas vraiment connue dans la mesure où une partie des diabétiques ignorent leur maladie et ne la traitent pas. Par ailleurs, il est délicat d'analyser l'accessibilité de ces malades aux prestations, puisqu'une partie des traitements peuvent être limités à des régimes alimentaires. Ces lacunes ne sont pas forcément graves, selon le type d'analyse projetée ; par exemple, l'analyse de la fréquence des contrô-

les ophtalmiques chez les diabétiques peut parfaitement être effectuée, même s'il ne s'agit que d'une partie d'entre eux (voir plus loin). Pour d'autres analyses, par exemple la justification de traitements neuroleptiques de patients psychotiques, la pertinence de l'indicateur pourrait être discutable parce que les patients sont en grande partie identifiés à partir de leurs traitements. Le test de faisabilité effectué sur la base des trois maladies sélectionnées – le cancer, le diabète et les maladies mentales – montre qu'un certain nombre d'indicateurs devraient pouvoir être proposés à condition d'être interprétés avec précaution. Il faudra cependant systématiquement s'interroger sur ces éventuelles limites, en confrontant le cas échéant les maladies détectées avec des données plus fiables (gold standard), de manière à documenter clairement la sensibilité et la spécificité des méthodes d'identification des maladies.

Un autre point essentiel est la validité de la mesure, c'est-à-dire l'absence de biais. Les données présentées ici sont partielles, parce qu'elles ne concernent que trois cantons et surtout qu'une partie de la population. Les assurés des caisses-maladie ne présentent pas tous les mêmes profils de risque et il serait imprudent d'inférer des conclusions générales de la présente étude. Par ailleurs, des différences entre cantons pourraient s'expliquer par des différences de pratique de facturation. Par exemple, le fait que des chimiothérapies anticancéreuses soient dispensées le plus souvent dans le cadre de traitements ambulatoires ou semi-hospitalier constitue par exemple un risque de biais non négligeable (informations manquantes). Enfin, la fiabilité des données et leur manque d'exhaustivité peuvent influencer sur les différences constatées.

La question de la précision est évidemment un point crucial dans les comparaisons. Les exemples présentés dans ce rapport sont assez souvent fondés sur des nombres relativement faibles de malades. Il est donc possible que les différences ne soient pas significatives d'un point de vue statistique.

Nous avons volontairement opté pour un ton démonstratif dans les exemples qui suivent, le but étant de susciter la curiosité et de montrer l'intérêt potentiel de ces comparaisons. Il serait toutefois abusif d'en tirer des conclusions générales sur les pratiques dans les différents cantons.

## 2 Etude de faisabilité

### 2.1 Indicateurs de couverture

Les indicateurs de couverture mettent en relation des volumes de prestations par rapport à une population générale. Par exemple, le taux de couverture du vaccin contre la grippe pourrait être mesuré en calculant la proportion de la population qui en a bénéficié ; ce taux pourrait être calculé en fonction du profil des sujets (personnes âgées de plus de 65 ans, hébergement médico-social) pour s'assurer que les directives de l'OFSP sont suivies. Une analyse par région pourrait le cas échéant mettre en évidence un besoin d'information des praticiens concernés.

Un autre exemple est fourni par les débats sur la clause du besoin. L'ordonnance du 3 juillet 2002 a en effet montré que décompter les médecins sur la base du registre des créanciers des assureurs maladie ne reflétait pas forcément la réalité, notamment parce qu'une partie non négligeable des médecins travaillent à temps partiel ou très partiel (retraités par exemple) ; par ailleurs, les numéros créanciers peuvent être actifs longtemps après l'arrêt d'activité en cas de contestation de factures non résolues. Le ratio entre le nombre de consultations et la population serait à cet égard un indicateur plus précis de la couverture des besoins pour déterminer la densité médicale.

Le taux de couverture médicale pourrait être analysé pour chaque discipline. Les données récoltées dans la présente étude permettent par exemple de calculer le rapport entre le nombre de consultations de psychiatres par rapport à la population souffrant de problèmes psychiatriques. La couverture est fournie par le nombre de consultations (prestations) par assuré (population), qui est varié fortement d'un canton à l'autre : 0,20 (17'593/87'882) en Argovie, 0,32 au Tessin et 0,45 dans le canton de Vaud (tableau 3, partie A).

Tableau 3 Couverture en consultations psychiatriques et autres indicateurs

<b>PSYCHIATRIE</b>	<b>Argovie</b>	<b>Tessin</b>	<b>Vaud</b>
Partie A. Couverture			
Nombre d'assurés (population)	87'882	32'884	49'226
Nombre de consultations (prestations)	17'593	10'393	22'257
Nb de consultations/assuré (couverture)	0.20	0.32	0.45
Partie B. Confrontation à d'autres indicateurs			
Autres descripteurs :			
Montants (ressources financières)	2'613'451	1'131'586	3'497'381
Nombre de patients psychiatriques (patients)	1935	1'263	1'703
Autres indicateurs :			
Nb de patients/assuré	2.2 %	3.8 %	3.5 %
Nb de consultations/patient (pertinence)	9	8	13
Coût/patient (adéquation)	1'351	896	2'054
Coût/consultation (productivité)	149	109	157

Il peut être intéressant de confronter ces taux aux autres indicateurs (tableau 3, partie B). La première question qui se pose est de savoir si la population argovienne souffre moins de problèmes psychiatriques que les autres. On y observe en effet une plus faible prévalence des problèmes psychiatriques traités (2.2 % en Argovie pour 3.8 % et 3.5 % dans les cantons de Tessin et de Vaud respectivement). Ce taux résulte à la fois de la morbidité (risque de souffrir d'un problème psychiatrique) et de l'accès aux prestations. Il n'est pas possible de distinguer ici ces deux notions, car le nombre de patients malades n'ayant pas consulté n'est pas connu<sup>1</sup>.

Il est intéressant de noter que la couverture en consultations psychiatriques est plus grande dans le canton de Vaud qu'au Tessin, parce que le nombre de consultations moyen par patient (pertinence) y est plus élevé (13 au lieu de 8). On pourrait dès lors se demander s'il n'y a pas une surconsommation de soins psychiatriques dans le canton de Vaud. A cet égard, il est intéressant de constater que le coût par consultation est presque 50 % plus élevé dans les cantons d'Argovie et de Vaud par rapport au Tessin. Cet écart pourrait être dû à un prix du point moins élevé<sup>2</sup>, mais aussi à une durée des consultations plus courte.

---

<sup>1</sup> On pourrait cependant l'estimer en se fondant sur les enquêtes sur l'état de santé de l'OFS.

<sup>2</sup> Le prix du point TarMed était plus faible en 2004 au Tessin, mais il s'agit ici de données 2002 pour lesquelles des comparaisons intercantionales étaient encore difficiles à établir faute d'un tarif fédéral.



## 2.2 Indicateurs de financement

Les indicateurs de financement mettent en relation les ressources consommées selon le profil de la population. Jusqu'ici, la population a été décrite en terme démographique, par classe d'âge et selon le genre. Les coûts réels peuvent ainsi être confrontés aux primes de l'assurance-maladie obligatoire, par caisse ou par région. L'enquête sur l'état de santé de la population permet en outre de connaître le montant des dépenses couvertes directement par les assurés (paiements directs, franchises, participation, etc.). Enfin, les dépenses de santé qui transitent par les impôts peuvent également être rapportées à la population.

L'intérêt du modèle présenté ici réside dans la possibilité d'affiner la description de la population en documentant également son état de santé. Outre l'intérêt sociologique de mesurer le degré de solidarité entre générations, régions ou autres groupes sociaux, la connaissance de l'état de santé de la population peut s'avérer utile pour la régulation de la LAMal. Rappelons que le risque financier encouru par les caisses-maladie dépend non seulement de l'âge et du sexe des assurés, mais également de leur état de santé.

Le montant du risque financier estimé par cette étude est de l'ordre 10'433 francs pour les cancers, 9'276 francs pour les affections mentales et 6'321 francs pour les diabètes, ces pathologies étant identifiées par la typologie SQLape® à partir de la seule statistique médicale des hôpitaux. Or les prescriptions de médicaments permettent d'identifier plus de cas : + 65 % de cas de cancer, + 375 % de cas de diabètes et + 362 % de cas de maladies mentales. Le surcoût estimé en tenant compte de toute l'information (statistique médicale et prescription médicamenteuse) est proche de celui estimé à partir des seules données hospitalières : 12'900 francs pour le cancer, 4'600 francs pour le diabète et 6'400 francs pour les affections mentales. La moindre sévérité des patients non hospitalisés explique les montants inférieurs du diabète et des affections mentales.

Il est important de noter que, quel que soit le modèle de financement ou d'assurance étudié (réassurance traditionnelle, développement d'un fonds de solidarité pour les hauts risques ou autres), il est nécessaire de disposer de données adéquates à la fois pour effectuer des simulations, calculer des primes ou encore corriger les éventuels effets pervers.

## 2.3 Indicateurs de morbidité

Les indicateurs de morbidité ont pour but de mesurer les risques de contracter une maladie. Certes, il n'est pas facile de détecter tous les cas de maladie sur la base des seules données disponibles de routine et il est probable qu'une partie des informations devront continuer à être récoltées par des enquêtes ad hoc (médecins sentinelles, déclaration obligatoire de certaines maladies très contagieuses). Toutefois, certains traitements médicamenteux permettront de mieux isoler certaines maladies dont la prévalence est aujourd'hui mal connue (asthmes, maladies inflammatoires au long cours, etc.) ; la seule connaissance du nombre d'emballages de médicament ne suffit pas, car il est nécessaire de pouvoir les rapporter à des patients précis.

Pour l'OFSP, ce type d'indicateur sert surtout à mesurer le risque de contracter des maladies qui peuvent être prévenues, ce qui n'est que partiellement le cas pour le diabète et les maladies mentales. Un suivi épidémiologique du risque de diabète insulino-dépendant serait parfaitement possible à partir des données récoltées, car la plupart des malades sont traités. En revanche, la prévalence du diabète non insulino-dépendant est probablement sous-estimée ici, certains de ces diabètes n'ayant pas toujours été diagnostiqués. De plus, une proportion importante des diabétiques diagnostiqués sont traités par le seul régime. En psychiatrie, les informations sur la morbidité à partir des médicaments sont limitées par les variations de doctrine de soins<sup>3</sup>.

L'intérêt de suivre les taux d'incidence du cancer est plus évident, mais il est alors indispensable de pouvoir distinguer les différents types de cancer (organe et malignité), si possible en documentant également le stade au moment du diagnostic (in situ, localisé, métastases ganglionnaires, autres métastases). Ceci n'est possible qu'à l'aide des registres du cancer, d'où l'intérêt de les associer à la démarche. Ces derniers pourraient également en retirer des avantages dans la mesure où la recherche scientifique pourrait s'appuyer sur les données disponibles pour mieux documenter le coût des traitements et le devenir des patients.

---

<sup>3</sup> Prescription de méthylphénidate pour le traitement de l'hyperactivité des adolescents, par exemple.

## 2.4 Indicateurs d'accessibilité

On distingue classiquement deux types d'indicateurs d'accessibilité, selon la distance que doit parcourir un malade pour recevoir des soins ou selon le délai d'attente. Mais il pourrait également être intéressant d'identifier d'autres types d'accès aux soins. La proximité des soins ne paraît pas être un problème aigu en Suisse, à l'exception de quelques régions reculées. Les délais d'attentes pourraient probablement être calculés à partir des données aujourd'hui disponibles, par exemple pour mesurer la durée des dialyses avant une transplantation rénale. Mais ceci ne pourrait être effectué qu'en intégrant les données des caisses-maladie et la statistique médicale des hôpitaux. Il faut cependant bien admettre qu'une connaissance plus systématique des délais d'attente restera limitée. En cas de besoin, des enquêtes de terrain pourraient être organisées, par exemple pour connaître les délais d'attente relatifs à des situations cliniques typiques (délai pour un rendez-vous d'un patient déprimé chez un psychiatre, temps d'attente pour la pose d'une prothèse de hanche, etc.).

L'analyse des données recensées dans la présente étude a pourtant mis en évidence des possibilités inattendues. On observe ainsi que les patients souffrant de cancer ont eu accès deux fois plus souvent aux services à domicile dans le canton de Vaud que dans le canton d'Argovie (tableau 4). Compte tenu de la durée limitée des prises en charge et du fait que cela permet aux patients concernés de rester dans leur entourage, cette question pourrait susciter quelques débats sur les différences de pratiques entre les cantons. On notera par exemple que les dépenses par prise en charge sont plus élevées dans le canton de Vaud, ce qui suggère une plus forte intensité des soins et peut-être le développement de soins spécialisés dans les soins palliatifs à domicile. Une analyse plus poussée permettrait probablement de voir si le prix de revient global est plus élevé, selon qu'il existe ou non un report de charges sur les hôpitaux par exemple.

Tableau 4 Accès des patients cancéreux aux services à domicile (Frs. = francs suisses)

Cantons	Argovie	Tessin	Vaud
Nombre de cancers	607	615	785
Accès aux services à domicile	84	117	222
Montants facturés par les services à domicile	Frs. 127'293	Frs. 168'618	Frs. 507'844
Accès aux services à domicile	14 %	19 %	28 %
Montant moyen par prise en charge	Frs. 1'515	Fs. 1'441	Frs. 2'287

On notera enfin, que les écarts s'amenuisent si l'on ne prend en considération que les prestations pour des accompagnements de fin de vie (décès dans l'année considérée) : 4.7 % en Argovie, 5.2 % au Tessin et 7.3 % dans le canton de Vaud. Ces taux sont sous-estimés parce qu'ils ne prennent pas en considération les décès survenant l'année 2003, alors que les prestations ont été prodigués à la fin de l'année 2002 par exemple. Des séries temporelles plus longues permettraient de tenir compte de la durée des soins à domicile.

## 2.5 Indicateurs de pertinence

La connaissance du profil des patients pris en charge permet de mieux évaluer la justification des prestations fournies. Il s'agit là d'une exigence clé de la LAMal (article 32). Les indicateurs de pertinence sont courants dans le domaine hospitalier. La mesure de la durée de séjour en est un exemple classique (nombre de journées par patient hospitalisé). Les taux de césarienne ou d'épisotomie (prestations) par hospitalisation pour accouchement (patient) sont des exemples souvent discutés. Dans le domaine médico-social, on pourrait citer le nombre de contentions (patients attachés dans leur lit) par pensionnaire d'EMS (patient) ou le nombre de toilettes effectuées de manière autonome par les patients. Le nombre de consultations par épisode de pneumonie serait un autre exemple dans le domaine ambulatoire.

L'application du modèle aux trois pathologies étudiées ici montre que les exemples concrets ne manquent pas, mettant souvent en évidence des différences intéressantes entre cantons. Par exemple, les patientes vaudoises ont reçu deux fois plus souvent des prescriptions de somnifères (tableau 5). Comment justifier la surconsommation dans le canton de Vaud et la différence entre hommes et femmes ?

Tableau 5 Consommation de somnifères parmi les patients psychiatriques (N=13'323)

	Argovie	Tessin	Vaud
Femmes	15 %	23 %	31 %
Hommes	11 %	22 %	24 %

Il est généralement recommandé de procéder à un traitement anti-oestrogénique chez les patientes de moins de 70 ans opérées du sein à risque modéré ou élevé de récurrence. Or, ces proportions semblent nettement trop faibles, même s'il faudrait confirmer ce résultat avec les données d'autres assureurs (tableau 6) et vérifier la fiabilité des relevés. En cas de confirmation, s'agit-il d'un problème de pratique médicale ou de mauvaise compliance des patients ?

Tableau 6 Fréquence des traitements anti-oestrogénique chez les patientes opérées du sein et de moins de 70 ans

	Argovie	Tessin	Vaud
Chirurgie du sein	18	15	18
Traitement anti-oestrogénique	8	1	7
Taux de traitement anti-oestrogénique	44 %	7%	39 %

La justification des prestations peut également être abordée dans le domaine de la prévention et de la promotion, par exemple pour le dépistage précoce des complications rétinienues du diabète. Alors qu'un contrôle ophtalmologique, est recommandé au moins une fois par an (parfois deux si les risques sont importants), seul 34 % y ont eu recours au Tessin, pour 36 % dans le canton de Vaud et 41 % en Argovie (tableau 7). Il sera intéressant de mettre ces proportions en rapport avec les taux de complications réellement observés sur une plus longue période.

Tableau 7 Proportion de diabétiques ayant consulté un ophtalmologue

Nombre de contacts	Argovie	Tessin	Vaud
Patients diabétiques	1'640	1'795	1'616
Ayant consulté un ophtalmologue	670	606	584
Taux	40.9 %	33.8 %	36.1 %

Un autre exemple d'actualité est lié à la consommation de médicaments génériques. Le taux de prescription de médicaments génériques en 2002 était ainsi différent d'un canton à un autre si l'on se réfère aux patients souffrant d'affections mentales (tableau 8).

Tableau 8 Proportion de médicaments génériques prescrits pour des patients psychiatriques

Nombre de prescriptions	Argovie	Tessin	Vaud
Médicaments génériques	3'350	1'165	5'118
Médicaments non génériques	59'328	92'946	113'354
Taux	5.6 %	1.2 %	4.5 %

Il pourrait également s'avérer intéressant d'analyser la diffusion des nouveaux médicaments. Ainsi, on observait en 2002 une plus forte proportion de prescriptions de médicaments autorisés depuis moins de 3 ans dans la liste des spécialités dans le canton d'Argovie que dans les deux autres cantons (tableau 9). Cette évolution est intéressante dans la mesure où chaque prescription coûtait en moyenne 161 francs cette année-là pour les nouveaux médicaments et 90 francs pour les plus anciens.

Tableau 9 Nouveaux médicaments (autorisés depuis &lt; 3 ans) prescrits par les psychiatres

Nombre de prescriptions	Argovie	Tessin	Vaud
Médicaments nouveaux	560	546	512
Médicaments non génériques	4'918	8'102	7'237
Taux	11.4 %	6.6 %	7.1 %

Il est évidemment intéressant d'aller au-delà de l'analyse des effets de la publicité pour analyser les pratiques médicales elles-mêmes. Il est ainsi possible de centrer l'analyse sur la prescription d'anti-dépresseurs par les psychiatres, par exemple (tableau 10). Les traitements par des inhibiteurs sélectifs de la récapitation de sérotonine (SSRI) sont les plus fréquents, conformément aux recommandations actuelles, qui les préconisent comme traitement de première intention, en raison de leur efficacité comparable aux antidépresseurs tricycliques (N06AA) avec moins d'effets indésirables cependant. On observe par ailleurs que les traitements combinant des SSRI et des antidépresseurs d'autres types soient en moyenne plus cher (+10 à +25 %) confirme le fait que ces traitements sont en général adressés à des patients souffrant de problèmes plus sévères ou résistants à un traitement par des SSRI<sup>4</sup>.

4 Il faudrait pour en être sûr également tenir compte des autres problèmes de santé qui ont également pu influencer sur les coûts totaux, ainsi qu'analyser le devenir de ces patients sur une durée assez longue pour se faire une idée de l'effet obtenu. On pourrait également vérifier que des antidépresseurs d'une même catégorie ne sont jamais prescrits simultanément.

Tableau 10 Antidépresseurs prescrits par les psychiatres (13'323 patients, année 2002)

Type de médication (codes ATC à 5 positions)		Argovie	Tessin	Vaud
N06AB	Inhibiteurs sélectifs de la récapture de la sérotonie (SSRI)	43 %	40 %	44 %
N06AX	Autres antidépresseurs	30 %	22 %	25 %
N06AA	Inhibiteurs non sélectifs de la récapture des monoamines	19 %	29 %	24 %
N06CA	Antidépresseurs combinés à des neuroleptiques	5 %	6 %	6 %
N06AG	Inhibiteurs de la monoamine oxydase A (IMAO)	1 %	1 %	0 %
N06BA	Sympathomimétiques centraux	1 %	2 %	1 %

Un autre exemple peut être tiré de l'analyse des traitements des patients diabétiques (tableau 11). Le traitement par l'insuline représente un tiers du coût médicamenteux total, celui des sulfonamides et des biguanides respectivement 24 % et 18 %, les thiazolidinediones 18 %, les autres médicaments 6 %. Le traitement par des thiazolidinediones est d'introduction récente, son coût plus élevé et son efficacité relative par rapport aux traitements classiques est discutée (cf. revue Prescrire 2005 ; 25(260) : 245-253). La fréquence observée dans les trois cantons paraît dans tous les cas supérieure à la prévalence des contre-indications aux traitements classiques. Là également, il est possible que les différences régionales soient plus dues à des efforts de marketing qu'à des raisons médicales. Il serait intéressant de pouvoir suivre l'évolution au cours du temps.

Tableau 11 Traitements médicamenteux du diabète (5'051 patients, année 2002)

	Argovie	Tessin	Vaud
Nombre de patients diabétiques	1'640	1'795	1616
Nombre de patients traités par des thiazolidinediones	134	287	209
Taux de traitement par des thiazolidinediones	8.2 %	16.0 %	12.9 %

## 2.6 Indicateurs d'adéquation

Les indicateurs d'adéquation mettent en relation les ressources consommées pour la prise en charge des patients. Dans le domaine hospitalier, il s'agit du coût par hospitalisation qui peut être analysé en fonction du profil des patients (maladies, interventions, etc.). Dans le domaine ambulatoire également, l'analyse des prix de revient par pathologies constitue un indicateur clé pour l'analyse du fonctionnement du système de soins. Or, l'attribution d'une prise en charge à une seule pathologie n'est souvent pas possible, car les patients présentent souvent des affections multiples. Attribuer l'ensemble des coûts à chaque maladie entraînerait une surestimation des coûts. Ce problème peut être illustré sur la base d'une analyse des coûts fondée sur seulement quatre maladies (tableau 12).

Tableau 12 Dépenses totales par maladie (milliers de francs)

	<b>Ambulatoire</b>	<b>Hospitalier</b>	<b>Global</b>
Tous les patients	289'188	102'261	391'450
Patients cancéreux	22'070	14'547	36'617
Patients diabétiques	35'498	15'583	51'082
Patients psychiatriques	65'635	36'228	101'864
Autres maladies	274'881	94'483	369'364
Somme des coûts des 4 maladies			558'928

Les patients cancéreux, diabétiques et psychiatriques ont été identifiés comme indiqué plus haut. Les autres maladies ont été déterminées à partir d'une liste de médicaments considérés comme forcément associés à une maladie significative.

Le coût total des patients pris en considération atteint la somme de 391 millions de francs, mais la somme des coûts des patients attribués à chaque maladie était de 559 millions de francs, soit plus de 50 % de plus. Ceci est dû au fait que les patients souffrent parfois simultanément de plusieurs maladies (tableau 13).

Tableau 13 Distribution des maladies

<b>Nombre d'assurés*</b>	<b>Autre maladie</b>	<b>Cancer</b>	<b>Diabète</b>	<b>Psychiatrie</b>
Autre maladie	97'062	–	–	–
Cancer	1'810	2'007	–	–
Diabète	4'837	240	5'052	–
Psychiatrie	8'363	428	867	9'222

\*Les nombres figurant dans la diagonale du tableau représentent les nombres d'assurés souffrant d'une seule maladie

Il est intéressant d'analyser les coûts moyens pour chaque association de maladie (tableau 14). On observe ainsi qu'un cancer coûte plus cher en moyenne par exemple s'il est associé à un diabète ou une affection psychiatrique que s'il est seul.

Tableau 14 Coûts moyens par maladie (francs)

Coût global*	Autre maladie	Cancer	Diabète	Psychiatrie
Autre maladie	3'805	18'446	9'886	11'077
Cancer	18'446	18'245	23'311	26'375
Diabète	9'886	23'311	10'111	18'276
Psychiatrie	11'077	26'375	18'276	11'046

\*Les montants figurant dans la diagonale du tableau représentent coûts moyens des assurés souffrant d'une seule maladie. On observe par ailleurs une symétrie des résultats par rapport à cette diagonale, le coût d'une association de deux maladies étant évidemment le même si on les intervertit.

Le surcoût de chaque maladie associée est fourni dans le tableau 15 ; il s'agit du coût supplémentaire d'une association de maladie par rapport à une maladie prise isolément<sup>5</sup>.

Tableau 15 Surcoûts moyens par maladie

Cout global	Autre maladie	Cancer	Diabète	Psychiatrie	Surcoût moyen
Autre maladie	0	202	-225	31	négligeable
Cancer	14'641	0	13'200	15'329	14'390
Diabète	6'081	5'066	0	7'230	6'126
Psychiatrie	7'271	8'130	8'165	0	7'855

On observe aisément que le surcoût associé à chacune des maladies est relativement stable : environ 14'400 francs pour un cancer (79 % du coût d'un cancer seul), environ 6'100 francs pour un diabète (61 % du coût du diabète seul) et environ 7'900 francs pour une affection psychiatrique (71 % du coût d'une affection psychiatrique isolée). On remarque également que les autres maladies n'influent pas le coût des trois maladies analysées ici. Ces observations permettraient ainsi de calculer facilement des coûts standard par année, tenant compte des multiples pathologies présentées par les patients. Il s'agit là d'un avantage majeur pour comparer les coûts induits par médecin traitant, en tenant compte du profil de leurs patients. Le contrôle fonderait ainsi sur des standards tenant compte de l'ensemble des assurés suisses, ce qui pourrait être très utile aux caisses-maladie.

Pour les médecins également, la comparaison des coûts pourrait être intéressante. Le tableau 16 montre par exemple que le coût des prises en charge des patients leucémiques varie considérablement d'un canton à l'autre. Le montant des honoraires médicaux est nettement plus élevé dans le canton de Vaud qu'en Argovie, de même que le coût des médicaments. Une interprétation possible est que les médecins vaudois utilisent des traitements plus récents et plus coûteux. Les nombres de cas sont toutefois insuffisants pour tirer des conclusions statistiquement significatives; il est par ailleurs nécessaire de distinguer les différents types de leucémies et lymphomes pour s'assurer de la comparabilité des patients, les mêmes thérapies pourraient être pratiquées selon les cantons en séjour hospitalier ou en ambulatoire. Cette validation devrait pouvoir être faite en confrontant les données des assureurs avec celles de registres du cancer.

5 Par exemple, le surcoût d'un diabète en présence d'un cancer représente en moyenne un montant de 5'066 francs (= 23'311 francs - 18'245 francs).



Tableau 16 Distribution du coût moyen des leucémies et lymphomes

<b>Leucémies</b>	<b>Argovie</b>	<b>Tessin</b>	<b>Vaud</b>
Nombre d'assurés	25	18	39
Hôpitaux	10'140	6'310	11'822
Laboratoires	380	290	663
Médecins	1'114	2'102	2'785
Pharmacies	1'780	3'222	4'233
Physiothérapeutes	83	118	226
Spitex	465	7	316
Etablissements médico-sociaux	0	0	1'648
Autres fournisseurs de soins	26	268	86
Coût total	13'989	12'316	21'779

## 2.7 Indicateurs d'effectivité

L'effectivité est le rapport entre les effets obtenus par la réalisation d'une prestation. Par exemple, un traitement antalgique sera considéré comme effectif s'il a réellement soulagé le patient de ses douleurs. De même, une opération de pontage coronarien sera considérée comme effective si elle a rétabli la circulation sanguine de l'artère opérée. La mesure de l'effectivité se distingue de celle de l'efficacité par le dénominateur qui décompte la quantité de prestations au lieu du nombre de prise en charges du patient. Par exemple, le taux d'infection nosocomiale post-opératoire est une mesure d'effectivité parce qu'il se rapporte à chaque événement à risque, soit pour chaque opération. Ce taux ne sera pas le même si on le mesurait en fonction du nombre de patients opérés, parce qu'un patient est susceptible de subir plusieurs opérations au cours d'une même hospitalisation.

Dans le domaine ambulatoire, il est possible de se faire une idée de l'effectivité des consultations en examinant le taux d'hospitalisations potentiellement évitables. Il s'agit d'hospitalisations liées à des complications qui ne devraient pas survenir si des soins adéquats sont délivrés à temps. Par exemple, les patients traités pour un diabète ne devraient normalement pas devoir être hospitalisés pour des complications<sup>6</sup> (coma diabétique par exemple). Or, ces taux semblent varier sensiblement d'un canton à l'autre avec un taux presque trois fois supérieur au Tessin par rapport aux deux autres cantons (tableau 17). L'attribution de l'hospitalisation à la qualité des soins ambulatoires doit néanmoins être discutée. En effet, les causes de variations peuvent être multiples : facteurs de risques diététiques, couverture en spécialistes du diabète ou en soins en matière d'éducation et d'information ou encore d'autres raisons à découvrir. Le taux d'hospitalisation plus faible dans le canton de Vaud est-il lié à un suivi ou un traitement plus intensif du diabète<sup>7</sup> ? Pourquoi le coût des complications est-il plus élevé dans le canton de Vaud que dans le canton d'Argovie, est-ce dû à une nature différente des complications ? Une analyse sur l'ensemble des cantons suisses serait intéressante pour y voir plus clair.

Tableau 17 Taux d'hospitalisation des patients pour diabète (N =5'052)

	Argovie	Tessin	Vaud
Nombre d'hospitalisations	33	93	31
Nombre de patients diabétiques	1'641	1'795	1'616
Taux	0.020	0.052	0.019
Coût des assurés non hospitalisés pour diabète	14'837	24'616	20'936
Coût des assurés hospitalisés pour diabète	20'653	37'211	34'370

<sup>6</sup> Seules les complications justifiant une hospitalisation sont prises en considération ici (catégorie SQLape® E-dD)

<sup>7</sup> Le coût sans complication est plus élevé dans le canton de Vaud qu'en Argovie, mais semblable à celui du Tessin.

## 2.8 Indicateurs d'efficacité

Les soins efficaces sont ceux qui permettent d'atteindre les résultats attendus avec un minimum d'effets secondaires indésirables. Ainsi, un traitement curatif est efficace s'il a favorisé la guérison du patient, un traitement palliatif s'il l'a soulagé et un traitement préventif s'il réduit les risques de contracter une maladie ou en diminue les conséquences. Le résultat attendu peut également être la pose d'un diagnostic ou une preuve de bonne santé (diagnostic écarté). L'efficacité peut se rapporter aux patients, par exemple en examinant la proportion de patients guéris ou le taux de survie ; elle peut également être mesurée de manière subjective par des enquêtes de satisfaction. Mais l'efficacité peut également être mesurée par les effets mesurés en fonction des malades (la létalité par exemple) ou la population générale (espérance de vie en bonne santé par exemple).

Pour le cancer, il sera possible de calculer des durées de survie en bonne santé lorsque des données seront disponibles sur plusieurs années. D'autres indicateurs devraient cependant pouvoir être calculés sur des intervalles de temps plus court. Ainsi, la proportion de femmes ayant bénéficié d'une chirurgie conservatrice en présence d'un cancer du sein<sup>8</sup> pourrait constituer un indicateur intéressant dans la mesure où l'on vise à minimiser les effets indésirables d'un traitement (ablation du sein, cf. tableau 18). Un taux élevé peut correspondre à un dépistage plus précoce ou à une prise de risque. L'analyse des durées de survie à ces cancers devrait en principe permettre de différencier les deux cas de figure<sup>9</sup>. Avec les données des registres du cancer actuellement disponibles, on pourrait imaginer restreindre l'analyse aux patientes opérées dans un délai de moins de trois mois après le diagnostic pathologique (biopsie).

Tableau 18 Proportion de cancers du sein opérés de manière conservatrice

	Argovie	Tessin	Vaud
Opérations conservatrices	16	11	13
Ablation du sein	8	19	34
Taux d'opération conservatrice	67 %	35 %	38 %

<sup>8</sup> Il s'agit des femmes ayant eu une opération du sein sans ablation (catégorie SQLape® MAM2).

<sup>9</sup> Nous ne l'avons pas fait ici, pour deux raisons : il faudrait disposer de séries temporelles plus longues et disposer des données des registres de cancer.

## 2.9 Indicateurs d'efficacité

L'efficacité est le rapport entre les effets obtenus et les ressources consommées pour les obtenir. Classiquement, on le définit comme le rapport entre le gain d'espérance de vie en bonne santé et le niveau de ressources consacrées au domaine de la santé (pourcentage du produit intérieur brut par exemple). Un tel indicateur est évidemment intéressant pour des comparaisons globales entre pays, mais il n'a pas de sens pour comparer l'efficacité des soins de différents médecins ou fournisseurs de soins. On pourrait par exemple calculer le gain d'espérance de vie obtenu par des soins curatifs, voire de pondérer ces années de vie gagnées par la qualité de vie pour des soins palliatifs. De tels calculs sont aujourd'hui difficiles à établir, car ils reposent sur des études ponctuelles coûteuses. La mise sur pied d'une base de données intégrant les données des statistiques hospitalières et des caisses-maladie permettrait de le faire de manière systématique à un coût bien moindre.

Dans le domaine ambulatoire, les résultats sont souvent moins spectaculaires mais néanmoins quantifiables. En se fondant sur les trois maladies ciblées par le présent rapport, on pourrait par exemple citer les indicateurs suivants :

- l'efficacité du traitement du diabète pourrait calculer en estimant l'impact des complications du diabète (coma diabétique, troubles rétinien ou vasculaires périphériques exprimés sous la forme d'un score par exemple) qui serait ensuite divisé par le coût du traitement (consultations et médicaments) ;
- l'efficacité du traitement du cancer serait estimée en divisant le nombre d'années de vie en bonne santé gagnées par le coût des prises en charge (investigations, traitement hospitalier et ambulatoire, etc.) ;
- l'efficacité des soins psychiatriques serait quant à elle évaluée en confrontant l'impact des soins (ampleur des symptômes résorbés grâce au traitement) à leur coût.

Pour aider le lecteur à se faire une idée à la fois de l'intérêt d'une mesure d'efficacité, mais aussi de sa difficulté, nous prendrons l'exemple de l'efficacité des soins prodigués aux patients souffrant de dépression. Les coûts étant connus, le problème est de savoir comment estimer l'impact des soins psychiatriques offerts à ces patients. L'idée générale est que l'évolution de la densité des contacts des patients avec les fournisseurs de soins reflète celle de leur état de santé. Il paraît en effet peu probable qu'un patient bénéficie de soins très coûteux s'il est en bonne santé et vice versa.

Pour illustrer la démarche, retenons les cinq patients suivants, qui ont tous coûté aux assurances au moins 1'000 francs au mois de janvier 2002 (tableau 19).

Tableau 19 Evolution du coût de prise en charge de cinq patients déprimés (exemple)

Patient	Jan.	Fév.	Mars	Avril	Mai	Juin	Juil.	Août	Sept.	Oct.	Nov.	Déc.	Score	Impact
Patient A	2'748	5'162	64	1'158	64	53	896	0	259	0	107	286	5	25'810
Patient B	1'639	4'865	15'968	15'347	13'311	8'250	7'320	1'518	111	1'035	1'540	1'341	4	63'872
Patient C	1'998	754	1'631	469	4'005	901	1'615	133	1'609	466	985	1'151	3	12'015
Patient D	2'460	2'162	2'321	2'464	2'220	2'256	2'509	2'220	2'200	2'546	2'239	2'259	2	5'092
Patient E	1'425	1'147	78	61	290	3'204	10'868	246	17	1'556	0	1'003	1	10'868

Chacun de ces cinq patients a présenté une dynamique propre. Le patient A a vu ses soins s'intensifier au mois de février, pour ensuite se réduire progressivement ; les frais observés depuis le mois d'août paraissant négligeables, on peut supposer que ce patient a guéri. Le patient B a également vu ses frais augmenter en début d'année, pour dépasser 15'000 francs par mois en mars et avril et décroître ensuite pour se stabiliser à environ 1'500 francs par mois ; on peut en inférer que le patient va sensiblement mieux mais qu'il n'est pas totalement guéri et qu'il a probablement quelques problèmes chroniques. Quant au patient C, sa situation ne s'est pas franchement améliorée même si les coûts du second semestre ont été un peu moins élevés. Le patient D souffre visiblement de problèmes chroniques qui nécessitent un traitement régulier et stable. Quant au patient E, sa situation est beaucoup moins stable, avec une aggravation au milieu de l'année. Certes, il s'agit là d'une vision réductrice de la trajectoire de chaque patient et il faudrait évidemment démontrer que l'évolution des coûts est corrélée à l'état de santé. Posons provisoirement cette hypothèse et tentons d'attribuer à tous les patients « déprimés » identifiés dans notre base de donnée un score estimant l'évolution de leur état de santé (tableau 20).

Tableau 20 Estimation de l'évolution de l'état de santé des patients déprimés (N=6'103)

Score	Evolution	Critères	Proportion
4	Guérison	Coût mensuel moyen de moins de 250 francs durant le dernier trimestre	17 %
3	Amélioration sensible	Coût du second semestre de moins de la moitié du premier semestre	16 %
2	Légère amélioration	Coût du second semestre inférieur à celui du premier, mais réduction <50%	33 %
1	Légère péjoration	Coût du second semestre supérieur à celui du premier, mais augmentation < 100%	26 %
0	Péjoration sensible	Coût du second semestre plus de deux fois plus élevé que celui du premier	7 %

Cette analyse montre que la psychiatrie n'est pas une discipline facile, puisqu'une petite partie des patients a guéri après plusieurs mois de traitement. Mais on constate quand même qu'environ deux tiers des patients vont mieux alors qu'un tiers ont vu leur état de santé se péjorer. On pourrait imaginer calculer l'efficacité entre totalisant les scores ainsi calculés et en les divisant par le coût total des patients ; l'examen du tableau 19 indique cependant qu'une telle vision attribuerait un meilleur résultat au patient A que pour le patient B, alors que ce dernier présentait visiblement des troubles plus sévères. Il a donc été décidé de mesurer l'impact d'une prise en charge en multipliant le score par la sévérité de l'affection. Faute de méthode éprouvée et disponible, nous pouvons estimer cette dernière par le coût mensuel maximal observé durant la période. Parmi les cinq patients du tableau 19, c'est le patient B qui aurait présenté les troubles les plus sévères, suivi du patient E. L'impact des soins serait dès lors maximal pour le patient B, suivi du patient A ; les résultats les moins bons seraient dès lors ceux du patient D. Cette évaluation paraît raisonnable dans la mesure où le problème psychiatrique du patient B était sévère et qu'il a été maîtrisé. A l'inverse, il semble correct de donner le plus mauvais score au patient D, car aucune amélioration n'a été notée alors que les coûts restent stables, voir augmentent au cours du temps.

Tableau 21 Estimation (grossière) de l'efficacité du traitement de la dépression (N=13'323)

Canton	Impact	Coût	Efficacité
Canton d'Argovie	7'123	7'016	1.02
Canton du Tessin	8'856	9'045	0.98
Canton de Vaud	8'227	9'640	0.85
Hommes < 40 ans	4'842	4'114	1.18
Femmes < 40 ans	5'604	4'486	1.25
Hommes 40–64 ans	5'453	5'559	0.98
Femmes 40–64 ans	5'912	5'757	1.02
Hommes 65 ans et plus	11'274	13'451	0.84
Femmes 65 ans et plus	11'625	13'412	0.87
Patients non suivis par un psychiatre	8'896	9'778	0.91
Patients suivis par un psychiatre	6'082	5'936	1.02

L'intérêt d'un indicateur dépend également de l'ampleur des questions qu'il met en lumière. Les résultats préliminaires sont à cet égard prometteurs, même s'il s'agit d'être prudent en raison des limites d'interprétation discutées plus loin. La comparaison de l'efficacité du traitement des dépressions entre cantons indique que l'impact est légèrement plus faible en Argovie, mais que ce canton aurait la meilleure efficacité. Le surplus de ressources consommées dans le canton de Vaud n'aurait pas les effets attendus selon cette analyse (15 % d'efficacité en moins).

L'analyse des résultats en fonction de l'âge et du genre des patients est également intéressante. De manière générale, l'impact des soins serait plus grand pour les personnes âgées, ce qui paraît logique étant donné que les symptômes y sont généralement plus sévères. Mais on observe que l'efficacité y serait également moindre, ce qui sous-entendrait qu'ils sont plus difficiles à résorber. Ceci paraît naturel dans la mesure où les perspectives sont souvent plus sombres pour les personnes âgées, ne serait-ce qu'en raison de leur état de santé général et d'une solitude plus marquée. On remarque également que l'efficacité est la meilleure dans la classe des jeunes femmes, ce qui pourrait s'expliquer par une évolution favorable plus forte des dépressions du post-partum.

Le but des indicateurs est de susciter des débats en vue d'améliorer le fonctionnement du système de santé. Une question est par exemple de savoir s'il vaut mieux pour un patient être suivi par un psychiatre ou par un médecin d'une autre spécialité (généraliste par exemple) en cas de dépression. Les résultats présentés dans le tableau 21 suggèrent une plus forte plus-value des psychiatres. On peut facilement imaginer les polémiques provoquées par la publication d'un tel classement par médecin. Il est donc essentiel de vérifier que la démarche proposée reflète la réalité. Les hypothèses posées ici sont en effet nombreuses et méritent d'être rappelées, notamment :

- les épisodes de dépression sont supposés tous commencer le 1er janvier 2002 ; or, l'évolution de l'état de santé devrait être mesurée après le début des symptômes, soit au début d'un épisode de maladie ; ceci suppose que la notion d'épisode de maladie soit clarifiée et que la période d'observation soit étendue sur deux ou trois ans ;
- les coûts pris en considération couvrent des prestations qui ne sont pas dévolues au traitement de la dépression ; il est nécessaire d'isoler le coût des prises en charge de la dépression de celles d'éventuelles autres maladies concomitantes ;

- l'évolution des coûts est supposée refléter celle de l'état de santé ; cette hypothèse devrait être vérifiée et il est possible que d'autres mesures plus appropriées, par exemple, on pourrait tenir compte des doses de médicaments prescrites pour se faire une idée plus précise de la durée des traitements (en divisant la dose cumulée par les doses journalières standard) et des classes thérapeutiques pour évaluer la sévérité des symptômes;
- la sévérité est évaluée à partir du coût mensuel maximal, ce qui est une approximation dont il faudrait vérifier qu'elle n'est pas biaisée ;
- les dépressions sont supposées être toutes équivalentes ; il serait probablement indiquer de différencier différentes situation : dépression du post-partum, dépression liée à un état de santé déficient, dépression associée à d'autres symptômes psychiatriques, etc. ;
- l'efficacité observée devrait être confrontée à l'efficacité attendue compte tenu des facteurs de confusion potentiels (au moins l'âge et le genre des patients) ;
- la manière de scorer les améliorations de l'état de santé est très réductrice ; il est recommandé d'analyser la sensibilité des résultats à d'autres manières de qualifier l'évolution des densités de contact ; il serait notamment intéressant d'adopter des outils topologiques plus adaptés et de les accompagner par des outils statistiques en vue de calculer des intervalles de confiance pour s'assurer que les variations ne sont pas dues au hasard.

Le but du présent chapitre n'est pas de convaincre le lecteur de la pertinence des indicateurs proposés. Il est évident que les choix proposés ici sont discutables et encore très provisoires.

### 3 Conclusions et recommandations

De manière générale, l'OFSP dispose des informations dont elle a besoin dans les domaines de la prévention, des soins hospitaliers et médico-sociaux. En revanche, elle manque cruellement d'indicateurs utiles pour analyser le fonctionnement du système de santé dans son ensemble. Par ailleurs, l'information des prises en charge ambulatoires est très lacunaire. La solution proposée ici est très attrayante parce qu'elle comble les principales lacunes, en proposant un cadre conceptuel complet, rigoureux susceptible d'apporter une certaine objectivité dans les débats au sujet de la LAMal et de la maîtrise des charges couvertes par la LAMal. En se limitant aux seules données aujourd'hui disponibles, elle permet par ailleurs de produire des indicateurs utiles à un coût tout à fait acceptable. Les enjeux sont importants et il est urgent de fournir à toutes les parties concernées (fournisseurs de prestation, assureurs, confédération, cantons, etc.) des outils plus modernes pour gérer le système de prestations de ce pays. Le cadre conceptuel préconisé par l'OFSP constitue une base minimale, de manière à concentrer un minimum de données et garantir ainsi des mesures efficaces de protection des données, empêchant l'identification de personnes physiques ou morales.

La présente étude a mis en évidence que la plupart de ces indicateurs ne pourraient être produits aujourd'hui faute d'un lien entre les bases de données existantes ou de concepts suffisamment clairs (épisodes par exemple), mais que les problèmes soulevés pouvaient être résolus pour la plupart :

- une batterie d'indicateurs en nombre limité permettrait de documenter l'ensemble des facettes de la performance du domaine des soins couverts par la LAMal;
- à l'exception des questions liés à l'accès aux prestations (temps d'attente), tous les indicateurs projetés pourraient être produits de routine, à condition qu'ils soient validés scientifiquement, et qu'un moyen soit trouvé pour intégrer les bases de données existantes ;
- la procédure du code de liaison anonyme mise au point par le service de cryptologie de l'armée du DDPS et actuellement utilisée par l'Office fédéral de la statistique constitue le seul moyen intéressant pour résoudre le problème; elle pourrait probablement être améliorée en cryptant également les numéros des assurés.

Cette étude a permis de montrer la sécurisation des données et également pourrait servir au cryptage des numéros des assurés l'opération de cryptage n'est pas une technique facile à maîtriser et il est très important de mieux en comprendre les fonctionnements ;

- des problèmes d'exhaustivité ont été observés, ils ont pu être analysés de manière approfondie pour pouvoir pallier aux déficiences constatées ; la question du circuit de l'information a été étudiée à ce titre : on pourrait par exemple imaginer que les données sur les consommations de médicaments soient fournies par les assureurs après validation ;
- les codes identifiant les fournisseurs de prestations ont pu être également anonymisés;
- la richesse de l'information recueillie va de pair avec des risques accrus en matière de protection des données ; des solutions efficaces de protection des données existent, mais elles sont contraignantes : des classes diagnostiques, opératoires et thérapeutiques devront être définies, des définitions claires devront être proposées pour définir les épisodes de soins, des mesures organisationnelles strictes devront être mises sur pied ; l'ensemble de ces mesures suppose que l'on définisse a priori quels sont les indicateurs prioritaires ;
- une instance neutre et soumise à des règles de contrôle claires devrait être définie pour intégrer les différentes bases de données existantes.



La démarche proposée a été testée avec succès sur trois maladies. Les exemples d'applications démontrent que les indicateurs pressentis ont pu être produits comme prévu, mais également que de nouveaux indicateurs propres à certaines maladies peuvent s'avérer très intéressants. La démarche est novatrice et la majorité des indicateurs devront encore être validés et peut-être affinés, mais il est important de noter que ces développements n'exigent pas de modifications quant aux données recueillies. L'opportunité d'étendre le recueil des données à d'autres sources (assurance invalidité, registres du cancer, registre des décès, etc.) pourront être décidée ultérieurement sur la base d'une analyse des avantages et inconvénients y relatifs ; dans tous les cas, les impératifs liés à la protection des données seront respectés.

La généralisation du modèle à tous les assureurs ne devrait pas poser de problème, dans la mesure où les données demandées sont les mêmes que celles précédemment fournies à l'OFSP dans le cadre des débats sur la révision de la LAMal (1997). L'extension aux données sur les médicaments ne devrait pas non plus poser de problèmes, puisque les assureurs disposent aujourd'hui déjà de la possibilité de recevoir l'ensemble des prescriptions délivrées par les pharmacies ; il s'agira simplement de s'assurer que les centrales de facturation des médecins fournissent les données dans les mêmes formats pour les médecins qui délivrent eux-mêmes les médicaments (Suisse allemande essentiellement). Il serait toutefois opportun de démarrer rapidement la mise en œuvre de cette statistique avec quelques assureurs pour résoudre les problèmes techniques ; l'appui du service de cryptage de l'armée du DDPS et du service de la protection des données de la confédération seront à cet égard très importants.

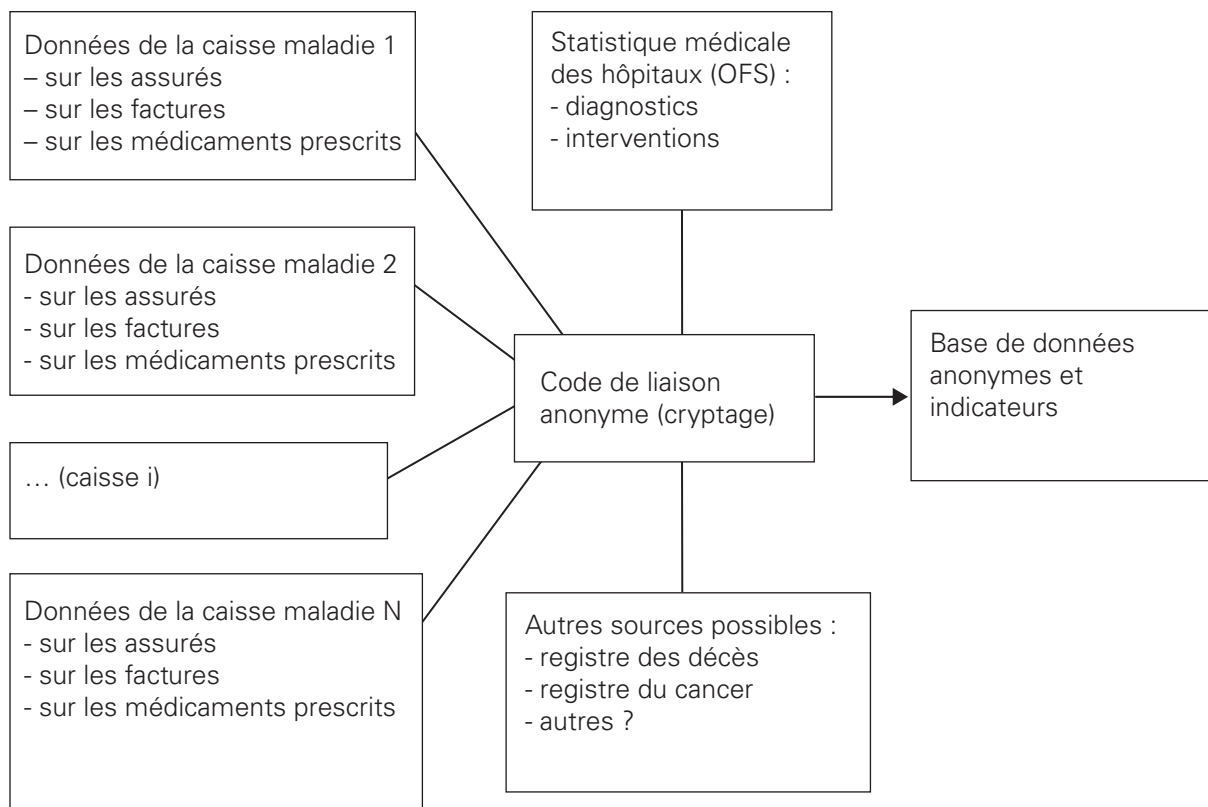
Hormis ces aspects liés au démarrage de ce travail, plusieurs problèmes scientifiques devraient être résolus pour permettre une généralisation du modèle à toutes les maladies. Ces problèmes sont listés dans le tableau 22.

Tableau 22 Travaux A réaliser

1. Modélisation des trajectoires de soins	2. Délimitation des épisodes ambulatoires	3. Identification des problèmes de santé	4. Etablissement de profils de pratique
	5. Détection des hospitalisations évitables avec des soins ambulatoires de qualité	6. Mesure de l'impact des soins sur l'état de santé	7. Mesure de l'incidence de maladies susceptibles d'être prévenues 8. Mesure de l'état de dépendance
9. Cryptage des codes de liaison anonymes & constitution de la base de données statistiques			

La réalisation de ces travaux ne sert pas que les objectifs de la LAMal. Les indicateurs proposés permettront également d'alimenter les réflexions des cercles de qualité (fournisseurs de prestations), d'améliorer l'analyse des pratiques médicales (formation continue pour les médecins, contrôle de gestion par les caisses-maladie) et de répartir les coûts par maladie (analyse coûts-efficacité des campagnes de prévention par exemple). On rappellera à cet égard que la réussite de ce travail passe par une intégration des données provenant de différentes sources (figure 2), ce qui suppose une anonymisation et une protection des données renforcée. Les avantages sont énormes, parce qu'ils permettent de mieux documenter les trajectoires de soins, de comparer les profils de pratique en tenant compte de l'état de santé des assurés et sans se restreindre à une seule caisse. Les mesures de protection des données vont bien au-delà de celles en vigueur aujourd'hui dans les caisses, ce qui est une condition cruciale pour la réussite de ce travail. Enfin, il ne faut pas négliger les impacts indirects sur la qualité et le coût des prises en charge, en disposant d'indicateurs mieux adaptés que ceux qui sont aujourd'hui disponibles.

Figure 2 Intégration des données



Le premier travail vise à modéliser les trajectoires de soins, en analysant les différentes configurations de prises en charge. Il s'agit notamment de définir dans quel cadre ont été effectuées les prestations :

- contrôles annuels (dépistages de cancer, contrôles oculaires, etc.)
- traitement d'affections chroniques (analyse des frais par année)
- soins de fin de vie (délimitation à définir)
- problèmes ponctuels (un contact isolé sans suite)
- épisodes ambulatoires (traitement d'une affection aiguë)
- épisodes mixtes (consultations précédent et suivant une hospitalisation)

Ce travail constitue un préalable à l'étape no 2, l'analyse par épisode de soins ne devant pas être entaché par des prises en charge ininterrompues (soins chroniques et de fin de vie) ou par des contacts uniques qui biaiserait les résultats. L'analyse des trajectoires de soins devrait par ailleurs clarifier le calcul de l'espérance de vie en bonne santé sans avoir besoin de recourir aux enquêtes de population, ce qui est un grand avantage pour l'analyse de l'impact des prises en charge par maladie<sup>1</sup>.

La seconde étape consiste à délimiter les épisodes de soins aigus, en définissant des délais entre les

<sup>1</sup> L'analyse peut en effet être effectuée avec une meilleure connaissance de l'histoire des maladies et avec une meilleure précision statistique (ensemble des assurés plutôt qu'un échantillon de personne).

épisodes dépendant des types de maladie du patient. Par exemple, deux prescriptions d'antibiotiques à 6 mois d'intervalle signifient qu'un patient a présenté deux épisodes d'infection, alors que la prescription d'anticancéreux dans le même intervalle se rapporte à un même épisode. La troisième étape vise à systématiser l'identification des maladies en se basant sur les informations de la statistique médicale des hôpitaux et des médicaments prescrits. Ceci devrait alors permettre d'établir les profils de pratique des médecins installés en cabinet, en tenant compte des affections présentées par les patients, ce qui n'est pas possible aujourd'hui. La détection des hospitalisations évitables vise à mesurer la qualité de la prise en charge d'affections qui ne devraient que rarement déboucher sur une complication justifiant une hospitalisation (asthme, coma diabétique, etc.) ; une validation scientifique de l'outil, basée sur des données médicales, sera cependant nécessaire pour s'assurer qu'il s'agit bien d'une conséquence des prises en charge. La mesure de l'impact des soins devrait en principe être ciblée sur le devenir des patients, pour déterminer si l'état de santé des assurés tend à s'améliorer, s'aggraver ou se stabiliser, en se fondant sur l'évolution de la densité des soins fournis.

L'étape no 7 vise à reprendre systématiquement les indicateurs actuellement utilisés en Suisse dans le domaine de la prévention, de manière à combiner au mieux les données existantes avec celles des assureurs. Quant à la mesure de l'état de dépendance des assurés, elle devrait prioritairement viser à déterminer si le patient est autonome d'un point de vue économique (AI) ou social (EMS) ; une bonne connaissance des statistiques de l'Assurance invalidité et des biais concernant l'hébergement des personnes âgées (services de maintien à domicile, sévérité des handicaps physiques et psychiques à l'admission en EMS, etc.).

L'importance et l'urgence des enjeux justifient que l'OFSP prenne quelques décisions rapidement, en vue de préparer l'avenir et doter la LAMal des instruments de pilotage qui lui manquent aujourd'hui :

1. Demander à quelques assureurs volontaires de livrer leurs données (cf. liste des variables de l'annexe A) pendant une phase de test.
2. Mettre en œuvre cette statistique avec l'appui du service de Cryptologie de l'armée DDPS et du service de la protection des données de la confédération
3. Une conception détaillée sera demandée à l'issue des travaux, avec les recommandations pratiques pour sa généralisation à tous les assureurs.

## 4 Utilisation des données mises à disposition

### A I Liste des données fournies par les assureurs maladie

Entités	Champs	Formats
<b>Contrat d'assurance</b>	– identifiant haché et crypté A	voir plus loin
	– assurance	nomenclature OFS
	– année de naissance (age de l'assuré au 1 <sup>er</sup> janvier l'exercice)	années
	– sexe	nomenclature OFS
	– canton de domicile	deux lettres
	– hospitalisation dans l'année	oui/non
	– séjour en EMS dans l'année	oui/non
	– frais hospitaliers à charge de l'assurance	francs
	– frais ambulatoires à charge de l'assurance	francs
	– participation financière de l'assuré	francs
	– montant de la franchise contractuelle	francs
	– début de validité de l'assurance	JJ.MM.AAAA
	– mois de fin de validité de l'assurance	JJ.MM.AAAA
	– décès éventuel durant l'année	0 = non, 1 = oui
<b>Factures</b>	– numéro de facture	nombre entier
	– identifiant haché et crypté A	voir plus loin
	– assurance	nomenclature OFS
	– séjour (hospitalier, EMS, autre)	1 caractère (H,E,A)
	– prescripteur	créancier santesuisse
	– prestataire	créancier santesuisse
	– date de début de période de facturation	JJ.MM.AAAA
	– date de fin de période de facturation	JJ.MM.AAAA
	– quantité de contacts avec des médecins	nombre entier
	– montant facturé	francs
<b>Médicaments</b>	– numéro de facture	nombre entier
	– pharmacode, code rayon	voir plus loin
	– quantité	nomenclature TarMed

## **A II Liste des données utilisées de la statistique médicale des hôpitaux**

Les enregistrements sont issus de la statistique médicale des hôpitaux pour les hospitalisations et semi hospitalisations terminées.

- rang du séjour durant l'année\*
- âge de naissance (age de l'assuré au 1<sup>er</sup> janvier de l'exercice)
- sexe
- canton de domicile
- diagnostics (tous)
- opérations (toutes)
- séjour avant l'admission
- séjour après la sortie
- durée de séjour\*

## **A III Médicaments marqueurs d'un cancer**

### **1. Marqueurs spécifiques de lymphomes et leucémies**

L01XB01	Procarbazine
L01XX14	Tretinoïn
L01BB02	Mercaptopurine
L01BB05	Fludarabine
L01DB06	Idarubicin
L01BB03	Tioguanine
L01BB04	Cladribine
L01BC01	Cytarabine
L01AA05	Chlorméthine
L01AA08	Prednimustine (également cancer du sein)
L01AA02	Chlorambucil (également cancer de l'ovaire)
L01XX05	Hydroxycarbamide
L01XX32	Bortezomib
L01XC02	Rituximab

Utilisation des données mises à disposition

L01XX01 Amsacrine  
L01XC04 Alemtuzumab  
L01AB01 Busulfan

## **2. Marqueurs spécifiques du cancer de la prostate**

L02AE01 Buserelin (chez l'homme uniquement)  
L02AE04 Triptorelin (chez l'homme uniquement)  
L02AE02 Leuprorelin (chez l'homme uniquement)  
L02AE03 Goserelin (chez l'homme uniquement)  
L02BB03 Bicalutamide  
L02BB02 Nilutamide  
L02BB01 Flutamide  
L02AA04 Fosfestrol  
L02AA02 Polyestradiol phosphate  
L01XX11 Estramustine

## **3. Marqueurs spécifiques du cancer du poumon (classé avec les autres cancers)**

L01BA04 Pemetrexed  
L01CA04 Vinorelbine (également le cancer du sein)  
L01XX34 Erlotinibe  
L01XX31 Gefitinib

## **4. Marqueurs spécifiques du cancer du sein**

L02BG06 Exemestane  
L02BG04 Letrozole  
L02BG03 Anastrozole  
L02BG02 Formestane  
L02BA02 Toremifene  
L02BA01 Tamoxifen

L02AB01	Megestrol (également le cancer de l'utérus)
L01CD02	Docetaxel
L01XC03	Trastuzumab
L02BA03	Fulvestrant

#### **5. Marqueurs spécifiques du cancer, mais non spécifiques d'une localisation mentionnée plus haut**

L01AA03	Melphalan
L01AA06	Ifosfamide
L01AC01	Thiotepa
L01AD02	Lomustine
L01AD06	Nimustine (cancer du cerveau et du rectum)
L01AX03	Temozolomide (cancer du système nerveux)
L01AX04	Dacarbazine
L01BA03	Raltitrexed (cancer digestif)
L01BC02	Fluorouracil
L01BC05	Gemcitabine (cancer du pancréas et des poumons)
L01BC06	Capecitabine (cancer digestif et du sein)
L01CA01	Vinblastine
L01CA02	Vincristine
L01CA03	Vindesine
L01CB01	Etoposide
L01CB02	Teniposide
L01CD01	Paclitaxel
L01DA01	Dactinomycin
L01DB01	Doxorubicin
L01DB02	Daunorubicin
L01DB03	Epirubicin
L01DB07	Mitoxantrone

Utilisation des données mises à disposition

L01DC01	Bleomycin
L01DC02	Plicamycin (cancer du testicule)
L01DC03	Mitomycin
L01XA01	Cisplatine
L01XA02	Carboplatine
L01XA03	Oxaliplatine (cancer digestif)
L01XC06	Cétuximab (cancer digestif et cutané)
L01XC07	Bévacizumab (cancer digestif)
L01XX17	Topotecan (cancer des ovaires et des poumons)
L01XX19	Irinotecan (cancer digestif)
L01XX28	Imatinib (leucémie et cancer digestif)
L02AB02	Médroxyprogestérone
L02AB03	Gestonorone
L03AC01	Aldesleukin (cancer du rein)
L03AX03	BCG vaccine (cancer de la vessie)
L03AX11	Tasonermin (cancer des muscles)
L03AX12	Melanoma vaccine

**6. Marqueurs de traitements adjuvants spécifiques au cancer  
(effets secondaires des traitements anti-cancéreux ou de métastases osseuses)**

M05BA02	Clodronic acid	Traitement de l'hypercalcémie (métastases osseuses)
V03AF05	Amifostine	Réduction de la toxicité des traitements (cyclophosphamide, cisplatine, radiothérapie)
M05BA08	Zoledronic acid	Traitement de l'hypercalcémie (métastases osseuses)
M05BA03	Pamidronic acid	Traitement de l'hypercalcémie (métastases osseuses)
V03AF07	Rasburicase	Traitement de l'hypercalcémie (métastases osseuses)

**7. Marqueurs de traitements adjuvants non spécifiques au cancer  
(effets secondaires des traitements anti-cancéreux ou de métastases osseuses)**

A04AA02	Granisetron	Vomissements liés aux cystostatiques (également : vomissements post-opératoires)
---------	-------------	---



A04AA03	Tropisetron	Vomissements liés aux cystostatiques (également : vomissements post-opératoires)
A04AA04	Dolasetron	Vomissements liés aux cystostatiques (également : vomissements post-opératoires)
H01CB02	Octreotide	Cancer digestif et du pancréas (également : acromégalie)
L03AA02	Filgrastim	Neutropénie (également : mobilisation des cellules souches avant leucophérèse)
L03AA03	Molgramostim	Neutropénie (également : causes non cancéreuses)
A04AA01	Ondansetron	Vomissements liés aux cystostatiques (également vomissements post-opératoires)
L03AA13	Pegfilgrastim	Neutropénie (également : causes non cancéreuses)
M05BA06	Ibandronic acid	Traitement de l'hypercalcémie (métastases osseuses, également : ostéoporose)
V03AF01	Mesna	Réduction de la toxicité des traitements (Endoxan, Holoxan)
V03AF03	Calcium folinate	Antidote méthotrexate, cancer du rectum, combinaison 5FU Anémie (pourrait être rendu spécifique par la DDD)
L03AA10	Lenograstim	Neutropénie (également : mobilisation des cellules souches avant leucophérèse)

### 8. Marqueurs non spécifiques du cancer (également indiqués pour d'autres maladies)

L01AA01	Cyclophosphamide	
L01BA01	Methotrexate	
L01XX33	Celecoxib	
L02BG01	Aminogluthetimide	
L03AA09	Sargramostim	
L03AA12	Ancestim	
L03AB04	Interferon alfa-2a	
L03AB05	Interferon alfa-2b	
L03AB06	Interferon alfa-n1	
L01XC05	Gemtuzumab	non commercialisé en Suisse
L01XX03	Altretamine	non commercialisé en Suisse

**9. Autres marqueurs spécifiques du cancer, mais non disponible en pharmacie en Suisse**

L01AA07	Trofosfamide
L01AB02	Treosulfan
L01AB03	Mannosulfan
L01AC02	Triaziquone
L01AC03	Carboquone
L01AD01	Carmustine
L01AD03	Semustine
L01AD04	Streptozocin
L01AD05	Fotemustine
L01AD07	Ranimustine
L01AG01	Etoglucid
L01AX01	Mitobronitol
L01AX02	Pipobroman
L01BC03	Tegafur
L01BC04	Carmofur
L01BC52	Fluorouracil, combinaisons
L01BC53	Tegafur, combinaisons
L01CC01	Demecolcine
L01CX01	Trabectedin
L01DB04	Aclarubicin
L01DB05	Zorubicin
L01DB08	Pirarubicin
L01DB09	Valrubicin
L01XC01	Edrecolomab
L01XD01	Porfimer sodium
L01XD04	Aminolevulinic acid
L01XD05	Temoporfin

L01XX02	Asparaginase
L01XX07	Lonidamine
L01XX08	Pentostatin
L01XX09	Miltefosine
L01XX10	Masoprocol
L01XX16	Mitoguazone
L01XX18	Tiazofurine
L01XX22	Alitretinoin
L01XX23	Mitotane
L01XX24	Pegaspargase
L01XX25	Bexarotene
L01XX27	Arsenic trioxide
L01XX29	Denileukin diftitox
L02AA01	Diethylstilbestrol
L02AA03	Ethinylestradiol
L02BG05	Vorozole
L02BG99	Inhibiteurs enzymatiques, inhibiteurs de l'aromatase (divers)
L02BX01	Abarelix

## **A IV Médicaments marqueurs d'un diabète**

### **1. Marqueurs spécifiques du diabète de type I**

A10AB01	Insulin (human)
A10AB02	Insulin (beef)
A10AB03	Insulin (pork)
A10AB04	Insulin lispro
A10AB05	Insulin aspart
A10AB30	Combinations
A10AC01	Insulin (human)
A10AC02	Insulin (beef)
A10AC03	Insulin (pork)
A10AC04	Insulin lispro
A10AC30	Combinations
A10AD01	Insulin (human)
A10AD02	Insulin (beef)
A10AD03	Insulin (pork)
A10AD04	Insulin lispro
A10AD05	Insulin aspart
A10AD30	Combinations
A10AE01	Insulin (human)
A10AE02	Insulin (beef)
A10AE03	Insulin (pork)
A10AE04	Insulin glargine
A10AE30	Combinations

### **2. Marqueurs spécifiques du diabète de type II**

A10BA01	Phenformin
A10BA02	Metformin

A10BA03	Buformin
A10BB01	Glibenclamide
A10BB02	Chlorpropamide
A10BB03	Tolbutamide
A10BB04	Glibornuride
A10BB05	Tolazamide
A10BB06	Carbutamide
A10BB07	Glipizide
A10BB08	Gliquidone
A10BB09	Gliclazide
A10BB10	Metahexamide
A10BB11	Glisoxepide
A10BB12	Glimepiride
A10BB31	Acetohexamide
A10BC01	Glymidine
A10BD01	Phenformin and sulfonamides
A10BD02	Metformin and sulfonamides
A10BD03	Metformin and rosiglitazone
A10BF03	Voglibose
A10BG01	Troglitazone
A10BG02	Rosiglitazone
A10BG03	Pioglitazone
A10BX02	Repaglinide
A10BX03	Nateglinide

### **3. Marqueurs non spécifiques du diabète de type II**

A10BF01	Acarbose	également intolérance au glucose
A10BF02	Miglitol	également intolérance au glucose

## **A V Médicaments marqueurs d'une maladie mentale**

### **1. Marqueurs spécifiques de troubles anxieux**

N05BA01	Diazepam
N05BA02	Chlordiazepoxide
N05BA03	Medazepam (dépression également)
N05BA04	Oxazepam (dépression également)
N05BA05	Potassium clorazepate (dépression également)
N05BA06	Lorazepam (dépression également)
N05BA08	Bromazepam
N05BA09	Clobazam (dépression également)
N05BA10	Ketazolam (dépression également)
N05BA11	Prazepam
N05BA12	Alprazolam
N05BA15	Camazepam
N05BA16	Nordazepam
N05BA22	Cloxazolam
N05BB01	Hydroxyzine (dépression également)
N05BC01	Meprobamate
N05BD01	Benzoctamine
N05BE01	Buspirone
N05BX01	Mephenoalone (dépression également)

### **2. Marqueurs spécifiques de la dépression**

N06AA01	Desipramine
N06AA02	Imipramine
N06AA04	Clomipramine
N06AA05	Opipramol (anxiété également)
N06AA06	Trimipramine

N06AA07	Lofepramine
N06AA08	Dibenzepin
N06AA09	Amitriptyline
N06AA10	Nortriptyline
N06AA12	Doxepin
N06AA14	Melitracen
N06AA16	Dosulepin
N06AA21	Maprotiline
N06AB03	Fluoxetine
N06AB04	Citalopram
N06AB05	Paroxetine
N06AB06	Sertraline
N06AB08	Fluvoxamine
N06AB10	Escitalopram
N06AF01	Isocarboxazid
N06AG02	Moclobemide
N06AX01	Oxitriptan
N06AX03	Mianserin
N06AX05	Trazodone
N06AX06	Nefazodone
N06AX11	Mirtazapine
N06AX16	Venlafaxine
N06AX18	Reboxetine
N06CA01	Amitriptyline and psycholeptics
N06CA02	Melitracen and psycholeptics

### **3. Marqueurs spécifiques de troubles psychotiques**

N05AA01	Chlorpromazine
N05AA02	Levomepromazine
N05AA03	Promazine
N05AA05	Triflupromazine
N05AB02	Fluphenazine
N05AB03	Perphenazine
N05AB06	Trifluoperazine
N05AB08	Thiopropazine
N05AC01	Periciazine
N05AC02	Thioridazine anxiété également
N05AC04	Pipotiazine
N05AD01	Haloperidol
N05AD04	Moperone
N05AD05	Pipamperone
N05AD09	Fluanisone
N05AE03	Sertindole
N05AF01	Flupentixol
N05AF03	Chlorprothixene (anxiété également)
N05AF05	Zuclopenthixol
N05AG01	Fluspirilene
N05AG02	Pimozide
N05AG03	Penfluridol
N05AH02	Clozapine
N05AH03	Olanzapine
N05AH04	Quetiapine
N05AL01	Sulpiride
N05AL03	Tiapride



N05AL04	Remoxipride
N05AL05	Amisulpride
N05AN01	Lithium (dépression bipolaire)
N05AX08	Risperidone
N05AX09	Clotiapine
N05AX12	Aripiprazole

#### **4. Marqueurs spécifiques de troubles du sommeil**

N05BA56	Lorazepam, combinaisons
N05CC01	Chloral hydrate
N05CD01	Flurazepam
N05CD02	Nitrazepam
N05CD03	Flunitrazepam
N05CD05	Triazolam
N05CD06	Lormetazepam
N05CD07	Temazepam
N05CD08	Midazolam
N05CD09	Brotizolam
N05CE02	Methyprylon
N05CF01	Zopiclone
N05CF02	Zolpidem
N05CF03	Zaleplon
N05CM01	Methaqualone
N05CM02	Clomethiazole
N05CM09	Valerian
N05CM13	Valnoctamide
N05CM15	Methylpentynol (dépression et anxiété également)
N05CX02	Methaqualone, combinaisons

## 5. Marqueurs spécifiques de problèmes psychiatriques non disponibles en pharmacie

N05BA07	Adinazolam	Psychiatrie - anxiété
N05BB02	Captodiame	Psychiatrie - anxiété
N05BX03	Etifoxine	Psychiatrie - anxiété
N05BX02	Gedocarnil	Psychiatrie - anxiété
N05BC51	Meprobamate, combinaisons	Psychiatrie - anxiété
N05BC04	Mebutamate	Psychiatrie - anxiété
N05BB51	Hydroxyzine, combinaisons	Psychiatrie - anxiété
N05BA23	Tofisopam	Psychiatrie - anxiété
N05BA21	Clotiazepam	Psychiatrie - anxiété
N05BA19	Etizolam	Psychiatrie - anxiété
N05BA18	Ethyl loflazepate	Psychiatrie - anxiété
N05BA17	Fludiazepam	Psychiatrie - anxiété
N05BA14	Pinazepam	Psychiatrie - anxiété
N05BC03	Emylcamate	Psychiatrie - anxiété
N05BA13	Halazepam	Psychiatrie - anxiété
N06AA13	Iprindole	Psychiatrie - dépression
N06AB07	Alaproclate	Psychiatrie - dépression
N06AB02	Zimeldine	Psychiatrie - dépression
N06AA23	Quinupramine	Psychiatrie - dépression
N06AA19	Amineptine	Psychiatrie - dépression
N06AA18	Dimetacrine	Psychiatrie - dépression
N06AF02	Nialamide	Psychiatrie - dépression
N06AA15	Butriptyline	Psychiatrie - dépression
N06AF03	Phenelzine	Psychiatrie - dépression
N06AA11	Protriptyline	Psychiatrie - dépression
N06AA03	Imipramine oxide	Psychiatrie - dépression
N06AA17	Amoxapine	Psychiatrie - dépression

N06AX07	Minaprine	Psychiatrie - dépression
N06AX19	Gepirone	Psychiatrie - dépression
N06AX17	Milnacipran	Psychiatrie - dépression
N06AX15	Pivagabine	Psychiatrie - dépression
N06AX14	Tianeptine	Psychiatrie - dépression
N06AX13	Medifoxamine	Psychiatrie - dépression
N06AX10	Oxaflozane	Psychiatrie - dépression
N06AB09	Etoiperidone	Psychiatrie - dépression
N06AX08	Bifemelane	Psychiatrie - dépression
N06AX04	Nomifensine	Psychiatrie - dépression
N06AX02	Tryptophan	Psychiatrie - dépression
N06AG03	Toloxatone	Psychiatrie - dépression
N06AF06	Iproclozide	Psychiatrie - dépression
N06AF05	Iproniazide	Psychiatrie - dépression
N06AF04	Tranlycypromine	Psychiatrie - dépression
N06AX09	Viloxazine	Psychiatrie - dépression
N05AA07	Chlorproethazine	Psychiatrie - psychose
N05AX11	Zotepine	Psychiatrie - psychose
N05AF04	Tiotixene	Psychiatrie - psychose
N05AE01	Oxypertine	Psychiatrie - psychose
N05AE02	Molindone	Psychiatrie - psychose
N05AA04	Acepromazine	Psychiatrie - psychose
N05AB01	Dixyrazine	Psychiatrie - psychose
N05AA06	Cyamemazine	Psychiatrie - psychose
N05AD08	Droperidol	Psychiatrie - psychose
N05AF02	Clopentixol	Psychiatrie - psychose
N05AH01	Loxapine	Psychiatrie - psychose
N05AK01	Tetrabenazine	Psychiatrie - psychose

Utilisation des données mises à disposition

N05AL02	Sultopride	Psychiatrie - psychose
N05AD02	Trifluoperidol	Psychiatrie - psychose
N05AD03	Melperone	Psychiatrie - psychose
N05AC03	Mesoridazine	Psychiatrie - psychose
N05AE04	Ziprasidone	Psychiatrie - psychose
N05AB05	Thiopropazate	Psychiatrie - psychose
N05AB09	Butaperazine	Psychiatrie - psychose
N05AX10	Mosapramine	Psychiatrie - psychose
N05AL06	Veralipride	Psychiatrie - psychose
N05AL07	Levosulpiride	Psychiatrie - psychose
N05AB07	Acetophenazine	Psychiatrie - psychose
N05AD06	Bromperidol	Psychiatrie - psychose
N05AX07	Prothipendyl	Psychiatrie - psychose
N05AD07	Benperidol	Psychiatrie - psychose
N05AB04	Prochlorperazine	Psychiatrie - psychose
N05AB10	Perazine	Psychiatrie - psychose
N05CX01	Meprobamate, combinaisons	Psychiatrie - somnifère
N05CX06	Dipiperonylaminoethanol, combinaisons	Psychiatrie - somnifère
N05CX05	Emepronium, combinaisons	Psychiatrie - somnifère
N05CX04	Clomethiazole, combinaisons	Psychiatrie - somnifère
N05CC04	Dichloralphenazone	Psychiatrie - somnifère
N05CM16	Niaprazine	Psychiatrie - somnifère
N05CX03	Methylpentynol, combinaisons	Psychiatrie - somnifère
N05CC02	Chloralodol	Psychiatrie - somnifère
N05CC05	Paraldehyde	Psychiatrie - somnifère
N05CD04	Estazolam	Psychiatrie - somnifère
N05CM03	Bromisoval	Psychiatrie - somnifère
N05CM12	Apronal	Psychiatrie - somnifère

N05CM11	Bromides	Psychiatrie - somnifère
N05CM10	Hexapropymate	Psychiatrie - somnifère
N05CM08	Ethchlorvynol	Psychiatrie - somnifère
N05CM07	Triclofos	Psychiatrie - somnifère
N05CM06	Propiomazine	Psychiatrie - somnifère
N05CC03	Acetylglycinamide chloral hydrate	Psychiatrie - somnifère
N05CM04	Carbromal	Psychiatrie - somnifère
N05CE03	Pyrrithyldione	Psychiatrie - somnifère
N05CE01	Glutethimide	Psychiatrie - somnifère
N05CD13	Cinolazepam	Psychiatrie - somnifère
N05CD12	Doxefazepam	Psychiatrie - somnifère
N05CD11	Loprazolam	Psychiatrie - somnifère
N05CD10	Quazepam	Psychiatrie - somnifère
N05CM05	Scopolamine	Psychiatrie – somnifère

**E I Prise en charge d'un assuré diabétique de 63 ans**

Date de début	Quantité	Montant	Prescripteur	Fournisseur	Prestations (codes ATC ou SQLape)
17.01.2002	1	78	3001 Médecin interniste	116 Pharmacies	A10AD Insulins and analogues, intermediate-acting combined with fast-acting
09.02.2002	3	235	3001 Médecin interniste	116 Pharmacies	A10AD Insulins and analogues, intermediate-acting combined with fast-acting
02.05.2002	2	156	3001 Médecin interniste	116 Pharmacies	A10AD Insulins and analogues, intermediate-acting combined with fast-acting
07.08.2002	1	3726	193 Hôpitaux	193 Hôpitaux	E-dD Diabète avec complication
09.08.2002	1	10	9546 Hôpitaux	116 Pharmacies	A12BA Potassium
09.08.2002	1	78	9546 Hôpitaux	116 Pharmacies	A10AD Insulins and analogues, intermediate-acting combined with fast-acting
09.08.2002	1	17	9546 Hôpitaux	116 Pharmacies	B01AC Platelet aggregation inhibitors excl. heparin
09.08.2002	1	41	9546 Hôpitaux	116 Pharmacies	C09AA ACE inhibitors, plain
22.08.2002	2	156	3001 Médecin interniste	116 Pharmacies	A10AD Insulins and analogues, intermediate-acting combined with fast-acting
26.08.2002	1	368	512 Hôpitaux	512 Hôpitaux	
28.08.2002	1	325	960 Hôpitaux	960 Hôpitaux	
29.08.2002	1	17	960 Hôpitaux	116 Pharmacies	H02AB Glucocorticoids
29.08.2002	1	43	960 Hôpitaux	116 Pharmacies	H02AB Glucocorticoids
29.08.2002	5	35	960 Hôpitaux	116 Pharmacies	N02AA Natural opium alkaloids
29.08.2002	1	90	960 Hôpitaux	116 Pharmacies	C09AA ACE inhibitors, plain
29.08.2002	2	37	960 Hôpitaux	116 Pharmacies	A06AD Osmotically acting laxatives
29.08.2002	1	4	960 Hôpitaux	116 Pharmacies	N02BE Anilides
29.08.2002	1	69	960 Hôpitaux	116 Pharmacies	A02BC Proton pump inhibitors
02.09.2002	1	711	859 Autre prescripteur	9467 Radiologue	
02.09.2002	4	207	960 Hôpitaux	960 Hôpitaux	
17.09.2002	7	283	388 Endocrinologue	388 Endocrinologue	
19.09.2002	1	297	388 Endocrinologue	662 Laboratoires	
03.10.2002	1	40	960 Hôpitaux	960 Hôpitaux	
08.10.2002	1	102	9546 Hôpitaux	9546 Hôpitaux	
11.10.2002	1	78	3001 Médecin interniste	116 Pharmacies	A10AD Insulins and analogues, intermediate-acting combined with fast-acting
14.10.2002	1	173	9546 Hôpitaux	9546 Hôpitaux	
15.10.2002	1	10	388 Endocrinologue	116 Pharmacies	N05CF Benzodiazepine related drugs
15.10.2002	1	79	388 Endocrinologue	116 Pharmacies	C10AA HMG CoA reductase inhibitors
15.10.2002	2	156	388 Endocrinologue	116 Pharmacies	A10AC Insulins and analogues, intermediate-acting
15.10.2002	2	194	388 Endocrinologue	116 Pharmacies	A10AB Insulins and analogues, fast-acting
15.10.2002	2	6	9546 Hôpitaux	116 Pharmacies	N02BE Anilides
15.10.2002	1	43	9546 Hôpitaux	116 Pharmacies	H02AB Glucocorticoids
15.10.2002	1	214	9546 Hôpitaux	116 Pharmacies	A02BC Proton pump inhibitors
22.11.2002	1	78	3001 Médecin interniste	116 Pharmacies	A10AC Insulins and analogues, intermediate-acting
22.11.2002	1	97	3001 Médecin interniste	116 Pharmacies	A10AB Insulins and analogues, fast-acting
27.11.2002	1	72	388 Endocrinologue	388 Endocrinologue	
29.11.2002	1	397	512 Hôpitaux	512 Hôpitaux	
16.12.2002	1	70	9546 Hôpitaux	9546 Hôpitaux	

**E II Prise en charge d'une assurée souffrant de problèmes psychiatriques**

Date de début	Quantité	Montant	Prescripteur	Fournisseur	Prestation (codes ATC ou SQLape)	
01.01.2002	1	227	Spitex	9111 Spitex		
01.01.2002	1	900	EMS	5994 EMS		
01.01.2002	1	2240	Autre	1912 Hôpital		
07.01.2002	1	9	Psychiatre	2001 Pharmacie	N05AA	Phenothiazines with aliphatic side-chain
07.01.2002	1	27	Médecins	2001 Pharmacie	N06AA	Non-selective monoamine reuptake inhibitors
07.01.2002	1	104	Médecins	2001 Pharmacie	N05AX	Other antipsychotics
19.01.2002	1	194	Autre	4872 Hôpital		
19.01.2002	1	300	Ambulance	10332 Ambulance		
22.01.2002	1	35	Médecin interniste	2001 Pharmacies	N05AA	Phenothiazines with aliphatic side-chain
22.01.2002	1	89	Médecin interniste	2001 Pharmacies	N06AA	Non-selective monoamine reuptake inhibitors
22.01.2002	1	257	Médecin interniste	2001 Pharmacies	N05AX	Other antipsychotics
22.01.2002	3	36	Médecin interniste	9664 Médecin interniste		
06.02.2002	1	3930	Médecin interniste	4872 Hôpital	P-fH	Psychose, hallucination et délire
15.02.2002	1	1416	Clinique psychiatrique	3241 Clinique psychiatrique		
07.03.2002	8	240	Médecin généraliste	9664 Médecin généraliste		
01.04.2002	1	2048	EMS	7619 EMS		
12.06.2002	4	162	Médecin généraliste	3287 Médecin généraliste		
01.07.2002	1	2070	EMS	7619 EMS		
18.09.2002	8	234	Médecin généraliste	3287 Médecin généraliste		
01.10.2002	1	2070	EMS	7619 EMS		
18.12.2002	1	42	Médecin généraliste	3287 Médecin généraliste		

**E III Prise en charge d'un assuré souffrant d'un cancer**

Date	Quantité	Montant	Prescripteur	Fournisseur	Prestation (codes ATC ou SQLape)	
09.01.2002	1	44	Gynécologue	4031 Laboratoires		
09.01.2002	1	97	Gynécologue	4031 Laboratoires		
09.01.2002	9	1361	Gynécologue	3824 Gynécologue		
11.01.2002	1	19	Hôpital	1130 Hôpital		
16.01.2002	1	22	Gynécologue	8901 Pharmacie	N05CF	Benzodiazepine related drugs
16.01.2002	1	32	Gynécologue	8901 Pharmacie	G03CA	Natural and semisynthetic estrogens, plain
17.01.2002	1	1362	Gynécologue	1130 Hôpital		
17.01.2002	1	218	Gynécologue	8554 Laboratoires		
25.01.2002	1	284	Hôpital	1130 Hôpital		
29.01.2002	1	549	Hôpital	1130 Hôpital		
30.01.2002	1	2460	Gynécologue	1130 Hôpital		
31.01.2002	1	344	Gynécologue	8554 Laboratoires		
14.02.2002	10	1329	Oncologue	2132 Oncologue		
01.03.2002	1	86	Gynécologue	4031 Laboratoires		
06.03.2002	1	318	Hôpital	1925 Hôpital		
09.03.2002	3	57	Oncologue	8901 Pharmacie	N05BA	Benzodiazepine derivatives
09.03.2002	1	13	Oncologue	8901 Pharmacie	A03FA	Propulsives
09.03.2002	1	200	Oncologue	8901 Pharmacie	A04AA	Serotonin (5HT3) antagonists
09.03.2002	2	14	Oncologue	8901 Pharmacie	N05BA	Benzodiazepine derivatives
09.03.2002	3	25	Oncologue	8901 Pharmacie	N05BA	Benzodiazepine derivatives
12.03.2002	1	2154	Oncologue	1260 Hôpital		
14.03.2002	1	202	Gynécologue	3824 Gynécologue		
02.04.2002	1	1876	Oncologue	1260 Hôpital		
16.04.2002	1	202	Gynécologue	3824 Gynécologue		
23.04.2002	1	1133	Oncologue	1260 Hôpital		
14.05.2002	1	1133	Hôpital	1260 Hôpital		
21.05.2002	5	609	Oncologue	2132 Oncologue		
29.05.2002	1	90	Radiologue	3115 Radiologue		
29.05.2002	2	120	Oncologue	8901 Pharmacie	D01AC	Imidazole and triazole derivatives
31.05.2002	1	485	Hôpital	1130 Hôpital		
03.06.2002	1	79	Gynécologue	4031 Laboratoires		
03.06.2002	2	265	Gynécologue	3824 Gynécologue		
06.06.2002	2	45	Gynécologue	8901 Pharmacie	N05CF	Benzodiazepine related drugs
07.06.2002	1	18	Gynécologue	8901 Pharmacie	N05BA	Benzodiazepine derivatives
07.06.2002	1	78	Gynécologue	8901 Pharmacie	A02BC	Proton pump inhibitors
10.06.2002	1	2790	Hôpital	9552 Hôpital	F-mM	Tumeur maligne du sein
11.06.2002	1	284	Gynécologue	8554 Laboratoires		
18.06.2002	1	23	Gynécologue	8901 Pharmacie	N05CD	Benzodiazepine derivatives
24.06.2002	7	5669	Radiologue	3115 Radiologue		
25.06.2002	1	40	Radio-oncologie	8901 Pharmacie	D03AX	Other cicatrizants
26.06.2002	1	22	Oncologue	8901 Pharmacie	N05CF	Benzodiazepine related drugs
26.06.2002	1	161	Oncologue	8901 Pharmacie	L02BA	Anti-estrogens
26.06.2002	2	6	Oncologue	8901 Pharmacie	N02BE	Anilides
26.06.2002	1	17	Oncologue	8901 Pharmacie	N05BA	Benzodiazepine derivatives



Date	Quantité	Montant	Prescripteur	Fournisseur	Prestation (codes ATC ou SQLape)	
28.06.2002	1	348	Radio-oncologie	1130 Hôpital		
29.06.2002	1	40	Radio-oncologie	8901 Pharmacie	D03AX	Other cicatrizants
02.07.2002	9	652	Gynécologue	7501 Physiothérapeute		
03.07.2002	1	17	Gynécologue	8901 Pharmacie	N05BA	Benzodiazepine derivatives
29.07.2002	1	167	Gynécologue	3824 Gynécologue		
20.08.2002	1	17	Oncologue	8901 Pharmacie	N05BA	Benzodiazepine derivatives
20.08.2002	1	22	Oncologue	8901 Pharmacie	N05CF	Benzodiazepine related drugs
20.08.2002	1	64	Hôpital	8901 Pharmacie	J02AC	Triazole derivatives
23.08.2002	1	37	Radiologue	8901 Pharmacie	A02BX	Other drugs for peptic ulcer and gastrooesophageal reflux disease (GORD)
27.08.2002	1	40	Radio-oncologie	8901 Pharmacie	D03AX	Other cicatrizants
27.08.2002	1	182	Radio-oncologie	8901 Pharmacie	J02AC	Triazole derivatives
11.09.2002	1	22	Radio-oncologie	8901 Pharmacie	N05CF	Benzodiazepine related drugs
11.09.2002	2	34	Radio-oncologie	8901 Pharmacie	N05BA	Benzodiazepine derivatives
11.09.2002	6	284	Radiologue	7501 Physiothérapeute		
17.09.2002	1	221	Oncologue	2132 Oncologue		
18.09.2002	2	150	Dermatologue	9196 Dermatologue		
19.09.2002	1	161	Oncologue	8901 Pharmacie	L02BA	Anti-estrogens
20.09.2002	1	162	Hôpital universitaire	9546 Hôpital universitaire		
28.09.2002	1	22	Oncologue	8901 Pharmacie	N05CF	Benzodiazepine related drugs
28.09.2002	1	17	Oncologue	8901 Pharmacie	N05BA	Benzodiazepine derivatives
30.09.2002	1	151	Radiologue	3115 Radiologue		
02.10.2002	1	164	Dermatologue	1859 Autres		
02.10.2002	1	95	Radio-oncologie	8901 Pharmacie	J02AC	Triazole derivatives
10.10.2002	2	139	Médecin interniste	9897 Médecin interniste		
11.10.2002	1	121	Médecin interniste	1260 Hôpital		
11.10.2002	2	45	Médecin interniste	8901 Pharmacie	N05CF	Benzodiazepine related drugs
11.10.2002	1	18	Médecin interniste	8901 Pharmacie		
11.10.2002	1	18	Médecin interniste	8901 Pharmacie	J07BB	Influenza vaccines
11.10.2002	1	18	Médecin interniste	8901 Pharmacie	J07BB	Influenza vaccines
11.10.2002	1	47	Médecin interniste	8901 Pharmacie	M01AH	Coxibs
11.10.2002	2	34	Médecin interniste	8901 Pharmacie	N05BA	Benzodiazepine derivatives
16.10.2002	9	415	Médecin interniste	1260 Hôpital		
31.10.2002	2	143	Ophtalmologue	1187 Ophtalmologue		
04.11.2002	1	200	Ophtalmologue	1806 Autres		
21.11.2002	1	17	Médecin interniste	8901 Pharmacie	N05BA	Benzodiazepine derivatives
21.11.2002	1	22	Médecin interniste	8901 Pharmacie	N05CF	Benzodiazepine related drugs
02.12.2002	1	62	Oncologue	6521 Laboratoires		
02.12.2002	1	266	Oncologue	2132 Oncologue		
03.12.2002	1	161	Oncologue	8901 Pharmacie	L02BA	Anti-estrogens
03.12.2002	1	17	Oncologue	8901 Pharmacie	N05BA	Benzodiazepine derivatives
03.12.2002	1	22	Oncologue	8901 Pharmacie	N05CF	Benzodiazepine related drugs
03.12.2002	1	47	Oncologue	8901 Pharmacie	M01AH	Coxibs
05.12.2002	1	3400	Oncologue	9552 Hôpital	MAM5	Reconstruction du sein

## Bibliographie

- Nr.: 1/98 Baur, Rita; Hunger, Wolfgang; Kämpf, Klaus; Stock, Johannes: Evaluation neuer Formen der Krankenversicherung. Synthesebericht  
Bestellen EDMZ\*318.010.1/98 d
- Nr.: 3/98 Baur, Rita; Eyett, Doris : Bewertung der ambulanten medizinischen Versorgung durch HMO-Versicherte und traditionell Versicherte. Untersuchungsbericht 3  
Bestellen EDMZ\*318.010.3/98 d
- Nr.: 4/98 Baur, Rita; Eyett, Doris: Selbstgetragene Gesundheitskosten. Untersuchungsbericht 3 (vergriffen)
- Nr.: 5/98 Baur, Rita; Ming, Armin; Stock, Johannes; Lang, Peter: Struktur, Verfahren und Kosten der HMO-Praxen. Untersuchungsbericht 4  
Bestellen EDMZ\*318.010.5/98
- Nr.: 6/98 Stock, Johannes; Baur, Rita; Lang, Peter; Conen, Dieter: Hypertonie-Management. Ein Praxisvergleich zwischen traditionellen Praxen und HMOs  
Bestellen EDMZ\*318.010.6/98
- Nr.: 7/98 Schütz, Stefan et al.: Neue Formen der Krankenversicherung: Versicherte, Leistungen, Prämien und Kosten. Ergebnisse der Administrativdatenuntersuchung, 1. Teil  
Bestellen EDMZ\*318.010.7/98
- Nr.: 8/98 Känzig, Herbert et al.: Neue Formen der Krankenversicherung: Alters- und Kostenverteilungen im Vergleich zu der traditionellen Versicherung. Ergebnisse der Administrativdatenuntersuchung, 2. Teil  
Bestellen EDMZ\*318.010.8/98 d
- Nr.: 9/98 Gabriel Sottas et al.: Données administratives de l'assurance-maladie : Analyse de qualité, statistique élémentaire et base pour les exploitations  
Bestellen EDMZ\*318.010.9/98 f

- 1 Eggli Y., Halfon P., Chikhi M., Bandi T. Ambulatory healthcare information system: A conceptual framework. *Health Policy* 2006; 78:26–38
- 2 Eggli Y., Halfon P., Chikhi M., Nguyen L., Decollogny A., Weissbaum F. Analyse des prestations prises en charge par la LAMal. Cadre conceptuel et étude de faisabilité centrée sur trois pathologies : cancer, diabète et affections mentales. Berne, OFSP, 2007.
- 3 Eggli Y., Halfon P., Bandi T., Chikhi M., Känzig H. Analyse des prestations prises en charge par la LAMal. Bern, Sécurité sociale, 2007 (accepted for publication, a copy is sent with the project).
- 4 International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10) : Instruction manual. World Health Organization : WHO Press, 1998.
- 5 Weiner S.L., Maxwell J.H., Sapolsky H.M., Dunn D.L., Hsiao W.C. Economic incentives and organizational realities: managing hospitals under DRGs. *Milbank Mem Fund Q* 1987; 4:463–487.
- 6 Halfon P., Eggli Y., Pretre-Rohrbach I., Meylan D., Marazzi A., Burnand B. Validation of the potentially avoidable hospital readmission rate as a routine indicator of the quality of hospital care. *Med Care* 2006;44(11):972-81.
- 7 Halfon P., Eggli Y., van Melle G., Chevalier J., Wasserfallen J.B., Burnand B. Measuring potentially avoidable hospital readmissions. *J Clin Epidemiol* 2002; 55:573–587.
- 8 Halfon P., Eggli Y., Prêtre-Rohrbach I., Meylan D., Marazzi A., Burnand B. Validation of the potentially avoidable hospital readmission rate as a routine indicator of the quality of hospital care. *Medical Care* 2006; 44(11); 972–981
- 9 Bagg W., Sathu A., Streat S., Braatvedt G.D. Diabetic ketoacidosis in adults at Auckland Hospital, 1988-1996. *N Z J Med* 1998; 28(5):604-8
- 10 Office fédéral de la statistique, section de la santé. Conception de plausibilité de la statistique médicale; version 4.0.1. 2005. [http://www.freudiger.com/medplaus/download/Plausikonzept\\_v401\\_20051027\\_public.pdf](http://www.freudiger.com/medplaus/download/Plausikonzept_v401_20051027_public.pdf)
- 12 Bouchard C. et al. Validation du codage des diagnostics et des interventions dans les établissements hospitaliers valaisans. Nice Computing 2005, Le Mont-sur-Lausanne. [http://www.ovs.ch/documents/validation\\_codage\\_2004\\_fr.pdf](http://www.ovs.ch/documents/validation_codage_2004_fr.pdf)
- 13 Billings J., Anderson G.M., Newman L.S. Recent findings on preventable hospitalizations. *Health Aff (Millwood)*. 1996; 15(3):239-49.
- 14 Guide to *Prevention Quality Indicators*. Agency for Healthcare Research and Quality, Rockville, MD. Available at [http://www.qualityindicators.ahrq.gov/pqi\\_download.htm](http://www.qualityindicators.ahrq.gov/pqi_download.htm)
- 15 Bindman A.B., Chattopadhyay A., Osmond D.H., Huen W., Bacchetti P. The Impact of Medicaid Managed Care on Hospitalizations for Ambulatory Care Sensitive Conditions. *Health Serv Res*. 2005; 40(1):19–38
- 16 Normand S.L.T., Zou K.H. Sample size considerations in observational health care quality studies *Statistics in Medicine* 2001; 31(3): 331–345
- 17 Hofer T.P., Hayward R.A., Greenfield S., Wagner E.H., Kaplan S.H., Manning W.G. The unreliability of individual physician “report cards” for assessing the costs and quality of care of a chronic disease. *JAMA* 1999; 281(22):2098–105.
- 18 OECD Health Data 2006: Statistics and Indicators for 30 Countries. [Database on CD-ROM] OECD Health Division. 2006. Available at [www.oecd.org](http://www.oecd.org)
- 19 Department of health. NHS performance indicators : February 2002. Available at [http://www.performance.doh.gov.uk/nhsperformanceindicators/2002/hacma\\_t.html](http://www.performance.doh.gov.uk/nhsperformanceindicators/2002/hacma_t.html). (accessed 28.02.2007)
- 20 Brown A.D., Goldacre M.J., Hicks N., Rourke J.T., McMurry R.Y., Brown J.D., Anderson G.M. Hospitalization for ambulatory care-sensitive conditions: a method for comparative access and quality studies using routinely collected statistics. *Can J Public Health* 2001; 92(2):155–9.

## Bibliographie

- 21 Sanderson C, Dixon J. Conditions for which onset or hospital admission is potentially preventable by timely and effective ambulatory care. *J Health Serv Res Policy* 2000;5(4):222-30.
- 22 Guide to *Prevention Quality Indicators*. Agency for Healthcare Research and Quality, Rockville, MD. Available at [http://www.qualityindicators.ahrq.gov/pqi\\_download.htm](http://www.qualityindicators.ahrq.gov/pqi_download.htm)
- 23 Millar J, Mattke S and the members of the OECD patient safety panel. *Selecting Indicators for Patient Safety at the Health System Level in OECD Countries*. OECD health technical papers. Paris, France: Directorate for Employment, Labour and Social Affairs, 2004
- 27 Falik M, Needleman J, Wells BL, Korb J. Ambulatory Care Sensitive Hospitalizations and Emergency Visits: Experiences of Medicaid Patients Using Federally Qualified Health Centers. *Medical Care* 2001;39(6):551-561.
- 24 Lebovitz HE, Austin MM, Blonde L, Davidson JA, Del Prato S, Gavin JR 3rd, Handelsman Y, Jellinger PS, Levy P, Riddle MC, Roberts VL, Siminerio LM; ACE/AACE Diabetes Recommendations Implementation Writing Committee. ACE/AACE consensus conference on the implementation of outpatient management of diabetes mellitus: consensus conference recommendations. *Endocr Pract*. 2006;12(Suppl 1):6-12.
- 25 Segreti J, House HR, Siegel RE. Principles of antibiotic treatment of community-acquired pneumonia in the outpatient setting. *Am J Med* 2005;118(Suppl 7A):21S-28S.
- 26 Ansari Z, Laditka JN, Laditka SB. Access to health care and hospitalization for ambulatory care sensitive conditions. *Med Care Res Rev*. 2006;63(6):719-41.
- 27 AHRQ. National Healthcare Quality Report . 2005. available at <http://www.ahrq.gov/qual/nhqr05/nhqr05.htm> accessed 28.02.2007
- 28 Homes P, Coles J, Goldacre M, Mason A, Wilkinson E. *Health Outcome Indicators: Diabetes*. Report of a working group to the department of health. Oxford: National Center for Health Outcome Development, 1999
- 29 Cameron AC, Trivedi PK. *Microeconometrics : methods and applications*. Cambridge University Press, 2005.
- 30 Brook RH, Chassin MR, Fink A, Solomon DH, Kosecoff J, Park RE. A method for the detailed assessment of the appropriateness of medical technologies. *International Journal of Technology Assessment in Health Care* 1986; 2: 53-63
- 31 Brown H, Prescott R. *Applied mixed models in medicine*. Chichester, John Wiley, 1999.
- 32 Cameron AC, Trivedi PK. *Regression analysis of count data*. Econometric society monograph no 30. Cambridge University Press, 1998.
- 33 Cameron AC, Trivedi PK. *Microeconometrics : methods and applications*. Cambridge University Press, 2005.

Rapports d'experts et de recherche pour l'assurance-maladie et accidents

# **Protection des données**

Privacy Preserving Data Collection  
Swiss Military, FU Br 41, KryptDet Report

Rainer Baumann, Marcel Rieser, Reto Strobl,  
Andrea Weisskopf, François Weissbaum

September 2005

## **Abstract**

Detailed data collection of individuals is a fundamental requirement for profound statistical analysis of a complex system as the health care system of Switzerland. The aim of this report is to model a privacy-preserving data collection system, to analyze an already proposed implementation and to develop a new proposal giving highest security.

# Contents

---

1	Introduction	75
2	Problem Statement and Background	76
3	Privacy Preserving Data Collection	77
3.1	System Model	77
3.2	Privacy Preserving Data Collection	79
3.3	An out-of-the-model Attack	81
4	Patient Identifier	82
5	A Scheme Providing Weak Anonymity	83
5.1	An Informal Overview	83
5.2	Verification of Requirements	85
6	A Scheme Providing Strong Anonymity	86
6.1	An Informal Overview	86
6.2	Technical Details	88
6.3	Verification of Requirements	90
6.4	Offline Recovery Authority	92
7	The Strong Scheme in Practice	93
7.1	System Architecture	93
7.2	Performance Estimations	94
8	Further Applications and Extensions	97
8.1	Medical Insurance Card	97
8.2	Payment and Controlling	97
8.3	Balance of Risk	97
9	Discussion and Conclusion	98
	References	99

---





# 1 Introduction

For understanding the uncontrolled growth of expenses of the Swiss health care system, a detailed privacy-preserving data collection is required. For this, we first present a system model and its requirements followed by two schemes implementing different levels of anonymity. Then we show a possible realization of a scheme in practice.

In our system model of a privacy-preserving data collection we distinguish the following entities.

**Sources** send their data to a **data collector** which stores them in a database. These data can be exported for statistical research by **data analysts**. In case of emergency a **recovery authority** is able to reveal an identity. For our system model we propose the following set of requirements: link-ability, anonymity, recoverability, integrity, authenticity, and completeness.

A scheme published in 1997 by the Swiss Federal Statistics Office has several flaws. It gives unlimited potency to the data collector because he has control over the encryption process as well as over the database. The scheme also lacks any authentication and exported data was not explicitly protected.

In this report, we propose a sophisticated scheme and a possible system architecture. Our scheme solves the mentioned flaws: guaranteeing anonymity by separating the recovery agency from the data collector, authenticating data over all stages, and implementing explicit data exporting employing privacy-preserving algorithms.

The rest of this report is structured as follows. In section 2 the problem statement and some background information are given. In section 3 we introduce our general model and define common notions. Following, in section 5 and 6 we present two possible schemes, one implementing weak and one strong anonymity. Then, in section 7, we give a possible realization of the scheme guarantying strong anonymity. In section 8 we present further applications and extensions. Finally the discussion and conclusion can be found in section 9.

## 2 Problem Statement and Background

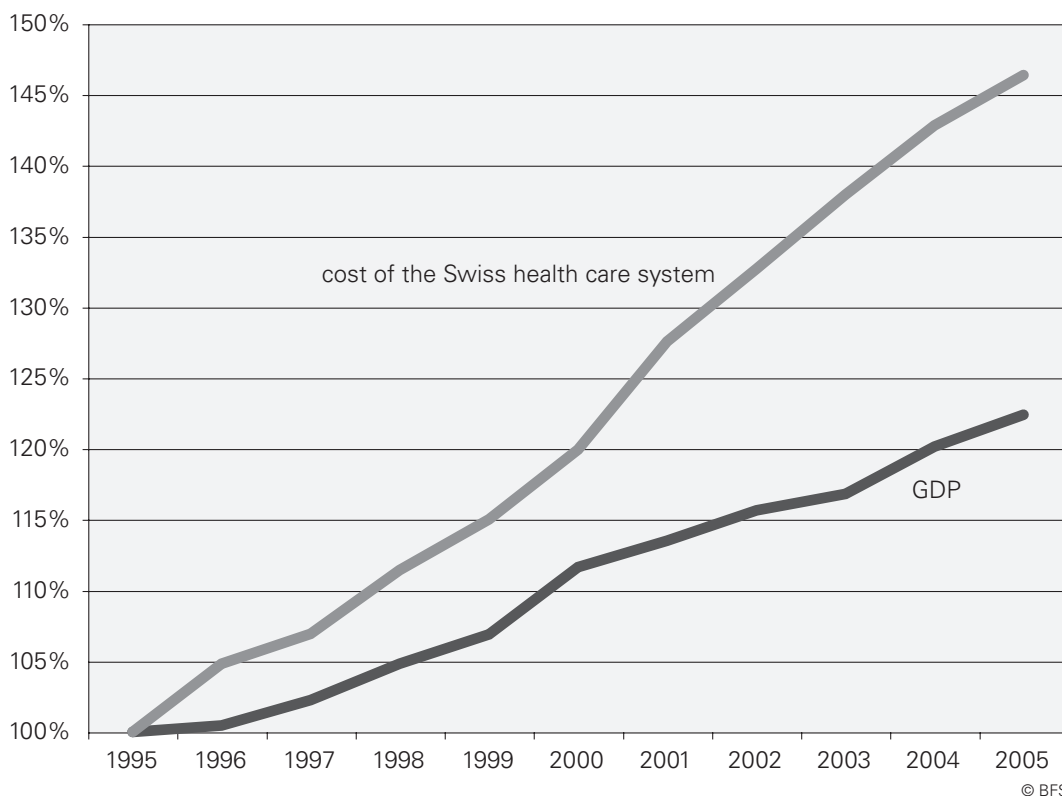
The Swiss people are proud of its excellent health care system. Unfortunately its cost runs out of control. In 1960 the costs added up to 5 % of the Swiss gross domestic product (GDP). A tremendous growth raised them up to 11.5% of the GDP or CHF 50 billiards in 2003. In the last decade this growth was even accelerated. Figure 1 shows the disparity of growth between the GDP and the expense for health care between 1995 and 2003.

To figure out the reasons of this unintentional growth, the Swiss Federal Statistics Office (SFS) collected statistical data in several areas of the health care system, mainly general profiles. These data allowed for creating a good overall view and in addition provided several detailed perceptions for several areas. But they do not allow to trace in detail treatments, their cost and their success for a certain disease. Based on the findings from the statistics, various different counter measures were proposed and realized. Unfortunately they did not make it to confine the continuous growth. According to an OECD analysis there are two major factors for the continuous growth: the large offer and the distortion of the price formation. For studying these effects and the cost efficiency of a treatment, it lacks a profound data collection with detailed traces about individuals. But collecting detailed data about individuals is very critical due to privacy concerns.

The aim of this report is to model and develop a privacy-preserving data collection system covering the whole health care system of Switzerland.

An additional, very critical requirement specified by the SFS for such a system is, that it must be possible to resolve the identity of a specific person if some findings from the data collection are very essential for this person. By definition, this requirement introduces some weakness into every possible model since it requires an instance which can break the whole system (for more details see section 3.3).

Figure 1 Development of cost of the Swiss health care system in percent compared to the GDP [2]



### 3 Privacy Preserving Data Collection

In this section, we explain what we mean by privacy-preserving data collection. In particular, we define the involved entities, the data-flow among these entities, and the properties that need to be guaranteed, as for example anonymity. Yet, we do not describe a specific implementation of such a system. This is the subject of following sections.

#### 3.1 System Model

Our system model consists of a set of **sources**, a central **data collector**, several **data analysts**, and of a central **recovery authority**. On a high-level, the relations among these entities can be summarized as follows. The sources are applications that provide the medical data to be collected. The data collector gathers this data from the sources, and makes it available to data analysts which perform the desired statistical data analysis. The system ensures that personal identifying data is replaced with pseudonyms once it leaves the sources, and that only the recovery authority is capable of linking a pseudonym to its respective identity.

Figure 2 illustrates the high-level overview of the various entities and their relation in our system model. It also labels the protocols that determine the data flow among the entities. In the following, we describe the components and protocols in more detail.

##### **Sources:**

The sources model client-applications that run at the institutes which generate the data to be collected. Such institutes could be for example hospitals, or insurance companies. We assume that every institute provides its data to its corresponding source application as a set of (non-anonymous) data records, each one representing a billable medical unit.

We assume that every data record is a tuple **(ID, data)** consisting of a privacy-critical part **ID** describing the personal identification information of the patient, and of actual payload data describing the medical unit (description, tax code, price, etc.). Deriving a unique identifier **ID** for a patient from personal identifying information is a hard task in itself. Section 4 proposes an algorithmic solution that should work well in practice. We remark that the payload may include personal identification information of the treating doctor, or the treating nurses, and so on. In practice, it may be desirable to protect their privacy as well. For simplicity, we do not address this explicitly here. Yet, we note that one can use the exact same techniques to make anonymous these identities as we use for the identity of the patient.

Notice also that in practice, it will be a tedious task to make all institutes provide their data in a common format, as it requires individual adjustments for every institute.

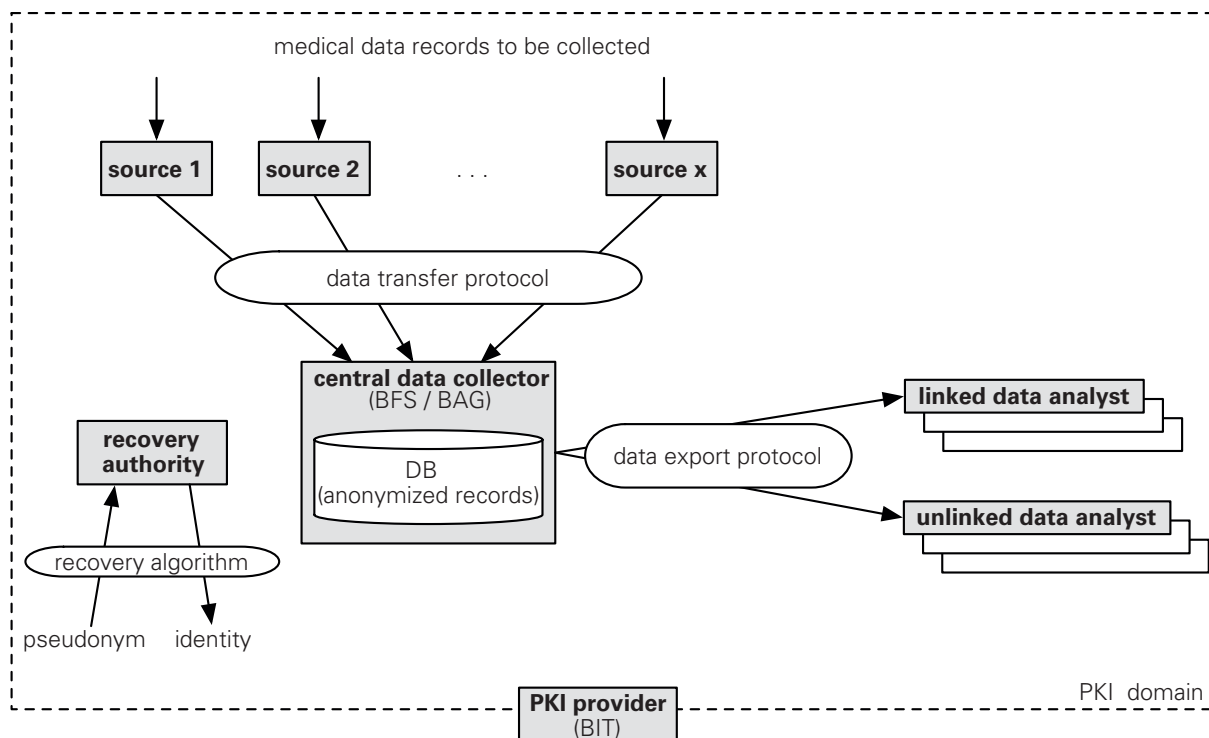
##### **Central Data Collector:**

The central data collector is the entity that gathers an anonymous version of the data records that the sources collect. In the anonymous records, the identifier **ID** is replaced by a random string **P**, which we also call a **pseudonym**. No entity but the recovery manager must be able to link a pseudonym **P** to its corresponding identity **ID**.

This transfer as well as the anonymization is achieved through a so-called **data transfer protocol** that is run between every source and the data collector. To ensure proper anonymization, the protocol may also involve the recovery authority.

In practice, the central data collector could be for example the Swiss Federal Statistics Office, the Federal Department of Defense, or the Swiss Federal Health Office.

Figure 2 Overview of the System Model. Grey boxes represent entities, ovals represent the protocols involved, and arrows represent the data flow.



**Data Analysts:**

Data analysts are third-parties such as universities, state departments, insurance companies, etc., which analyze the collected information. We assume that there is a well-defined process on how data analysts can request a set of records from the data collector, and a law defining when such requests are valid and may be served. The data collector is responsible to ensure the enforcement of these laws. Once it has decided to serve a request, it transfers the requested data records to the data analyst through a so-called **data export protocol**. During this transfer, a **pseudonym P** may either stay the same, or be replaced with a **'fresh' pseudonym P'**, or may be omitted completely.

We call a data analyst that receives records **with** pseudonyms a **linked-data analyst**, as he can reconstruct the medical history of a particular patient. On the other hand, we call a data analyst that receives records without pseudonyms an **unlinked-data analyst**.

**Recovery Authority:**

There may be situations, where the identity of a certain patient (or doctor, etc.) has to be recovered from a pseudonym. In our model, this can be done by a so-called **recovery algorithm**, which is run by the recovery authority. On input of a pseudonym **P**, the algorithm outputs the respective identity **ID**. We assume that in practice, there is law defining the circumstances under which such a recovery is legitimate. One has to trust the recovery authority that it only conducts legitimate recoveries.

It is crucial for the security of the system that one builds the recovery authority in a high-security environment, with limited access by people and programs.

In practice, it may also be desirable to have an independent entity that is responsible for the invocation of the recovery algorithm, and which keeps track of whom recovered whom on a public list. This is to make the anonymity as transparent as possible to the affected individuals.

**Public Key Infrastructure:**

We assume that there is a public key infrastructure (PKI) that allows us to authenticate sources, the data collector, the recovery authority, and the data analysts.

In practice, such a PKI could be managed for example by the Swiss Federal Office of Information Technology (BIT). The involved keys could be distributed by smart-cards, or secure USB sticks.

**The Network:**

We assume that all entities can communicate with the central data collector through authentic reliable **links**. Specifically, if some entity sends a message ***m*** to the data collector, then the network guarantees that the data collector eventually receives ***m***.

In practice, one can establish such a network by communicating over SSL/TLS connections with client and server authentication. The necessary certificates and private keys could be stored for example on the smart-cards used by the PKI.

**3.2 Privacy Preserving Data Collection**

The system model so far describes the general working of a system for privacy-preserving data collection. Yet, it does not specify the concrete implementation of the basic components, which are the ***data transfer protocol***, the ***data export protocol***, and the ***recovery algorithm***.

We call a concrete implementation of these components a privacy-preserving data collection scheme.

Having this abstract notion of a scheme allows us to describe the desired properties of such a ***scheme*** in an implementation-independent way.

In the following, we discuss informally these properties of a privacy-preserving data collection ***scheme***. This will serve as the basis for the analysis of the proposed solutions.

**Link-ability:**

For statistical purposes, it is desired that the data collector can link pseudonyms that originated from the same patient to each other (without knowing the actual identity of the patient). We can express this requirement in terms of the following relation between the records that the sources send and the records that the data collector stores in its data base:

Consider two sources, each one transferring a record ***(ID, data)*** and ***(ID', data')***, respectively, to the data collector. If ***ID*** and ***ID'*** are the same, then the data collector will store both records under the same pseudonym.

We should keep in mind that link-ability induces an anonymity problem: If the link between a pseudonym and an identity is discovered once, then the complete history of this patient is revealed.

One particular example where this introduces bad side-effects is a source and a data analyst that pool their information. By matching records of the analyst with records of the source using the message payload, one can derive pseudonym to identity relations, and thereby, recover the personal medical history of any patient treated by this source to the horizon of the data hold by the analyst.

This is bad because a source should only be responsible for the anonymity of the records that it collects itself, and not about records collected by other sources. Link-ability breaks with this natural requirement: If a source fails to keep secret the data of one of its patient, then not only the data of this source about this patient will lose its anonymity, but also the data collected by other sources on this patient.

### **Anonymity:**

There are two natural levels of anonymity a patient would expect: On the first level, he wants to remain anonymous with respect to data analysts, i.e., a data analyst must not be able to deduce a patient identity given only some records where patients are anonymous. On the second level, the patient wants to remain anonymous with respect to the data collectors. This is to avoid a big-brother scenario, where the central data collector (i.e., the state) keeps files for individuals. We summarize these properties below as weak and strong anonymity, respectively:

Consider two patients  $ID$  and  $ID'$  both treated at sources  $S = \{S_1, \dots, S_j\}$ , and let  $P$  and  $P'$  denote their respective pseudonyms.

### **Weak anonymity:**

If all sources in  $S$  and the data collector follow the protocol, then any coalition of sources not included in  $S$  and data analysts can distinguish  $(ID, P)$ ,  $(ID', P')$  from  $(ID, P')$ ,  $(ID, P)$  with a probability greater than  $1/2$ .

### **Strong anonymity:**

If all sources in  $S$  follow the protocol, then no coalition of the **data collector**, sources not included in  $S$ , and the data analysts can distinguish  $(ID, P)$ ,  $(ID', P')$  from  $(ID, P')$ ,  $(ID, P)$  with a probability greater than  $1/2$ .

Notice that strong anonymity implies weak anonymity, as the data analysts get their data from the data collector. Notice also that the above definitions already take link-ability into account. Specifically, they only require anonymity of a patient if all sources that treated this patient keep their data secret. If link-ability was not a requirement, we could strengthen the above requirements in the sense that the anonymity of the data collected by a certain source is provided as long as the collecting source keeps the data secret.

### **Recoverability:**

In certain cases, it may be desirable to recover the identity of a patient from its pseudonym. This could be the case, for example, if a data analyst discovers a pattern indicating a disease that requires immediate treatment. In this case, it would be in the interest of the patient to trade his anonymity for health. In other words:

Let  $(P, \text{data})$  denote a record held by either the data collector or a data analyst, and suppose the pseudonym  $P$  originated from a patient  $ID$ . If the recovery authority invokes the recovery algorithm with input  $P$ , then it outputs  $ID$ .

### **Authenticity and Integrity**

A fundamental requirement of a data collection scheme is that the data collector and the data analyst receive authentic records without modification, i.e., records have been inserted by an accredited source and have not been modified.

We formulate this in terms of the following relation between records that sources send and the records that the data collector (or data analysts) receive.

If the data collector (or a data analyst) receives a record  $(P, \mathbf{data})$  from the data transfer protocol (or the data export protocol, respectively), then the following holds:

1. If the recovery authority invokes the recovery algorithm, it outputs an existing identity  $ID$
2. There exist a source  $S$  that used the record  $(ID, \mathbf{data})$  as input to the data transfer protocol.
3. The  $\mathbf{data}$  contained in the data collector are exactly the same as the  $\mathbf{data}$  introduced by the source  $S$ .

Notice that integrity only guarantees that every record in the database has been provided by some source. Yet, it does not guarantee that every record represents a medical treatment unit that actually has occurred.

### 3.3 An out-of-the-model Attack

The requirements given in the previous section provide a comprehensive protection of the patient's anonymity. Yet, even if a scheme satisfies all these requirements, there is still an attack that is not captured by these security properties.

The problem is that one cannot verify if the records provided by a source represent medical treatment units that actually have occurred. In other words, a source could invent records, which would then end up in the database just like the genuine records.

To see how this leads to a problem, suppose a source  $S$  invents a record  $(ID, \mathbf{data})$  for a patient  $ID$ . According to our anonymity requirement, this source should not be able to learn the pseudonym(s) of this patient, even if it colludes with the data analysts or the data collector. However, since the invented record will end up in the database just like any other genuine record,  $S$  can compare the payload  $\mathbf{data}$  of the invented record with the payload of all records in the database. If  $S$  chooses the payload uniquely enough, it will match the payload of only a small set of records from the data collector's database. This gives only a small set of possible pseudonyms for the identity  $ID$ . Repeating this step with another record will reduce the number of candidate pseudonyms gradually to one.

In summary, as long as we do not have a notion of genuine data records, and a way to verify this property, no data collection scheme can ensure either weak or strong anonymity.

A natural definition of a *genuine record* could require that the record must be signed by the patient, i.e., by the individual whose anonymity could be compromised. This is only feasible if every patient can be authenticated by our PKI. In practice, this means that every patient needs a smartcard (or a similar hardware token) issued by the PKI provider that is capable of generating a digital signature. Currently, this is clearly not the case in practice. However, this will change with the introduction of the Swiss medical card (see Section 8). We suggest to reconsider this approach when this card is introduced.

Another way to increase the probability to have only genuine records is to sign every record by one or more individuals. Such individuals can be seen as 'data security officers' of a given source, and could be held responsible for forgeries.

For the purpose of this work, we ignore this problem and assume that all records are genuine.

## 4 Patient Identifier

In this section we discuss the necessary properties of a unique identifier and several possible implementations. They show that a medical card number would be the best choice but that a combination of common personal attributes is also a reasonable possibility.

As introduced in section 3.1, the privacy-preserving data collection system has to deal with many different data sources. For combining them to a consistent overall picture, it is necessary that all data sets of a person include the same unique identifier, an **ID**. There are several possibilities how to create such an **ID** as for example:

1. Social security number
2. Combination of common attributes as first name, family name, date of birth, gender
3. Medical card number

Before discussing the listed possibilities, let us quickly define several requirements for the identifier. It should be

1. Unique.
2. Available for everybody (neonate, foreigner with or without permanent residence, ...)
3. Constant over the whole life of an individual.
4. Not obviously contain personal data.
5. Known by every data source.

For example, the postal address is not suitable since it contravenes at minimum with 2 to 5. At a glance, the social security number seems to be a quiet good choice. But when we check it with our requirements, it infringes upon number 2, 4 and 5. A combination of common attributes has also some problems. It violates the requirements 1, 3 and 4. But the violation of 4 can be solved by applying a cryptographic hash function as described in [5] and the probability for a violation of 1 can be kept quiet low [1]. If well designed, a medical card number is the perfect solution since it fulfills all requirements but it is not yet available.

Recapitulating, a medical card number would be the best solution but it is not yet available. So we propose for the time being to use a combination of common attributes as described in [5] but to switch as soon as possible to a medical card number.

To get an idea on how unique **ID**, we briefly summarize an experimental study [1]. When calculating **ID** for over 220'000 individuals, 0.3% false-positives were found. This means that 0.3% of the data would not reflect the medical history of real individuals, but the combined history of at least 0.6% of all individuals, if we consider that each false-positive record consists of the data of two persons. If sometimes three or more individuals' data are merged because of the same **ID**, the percentage will be even higher. With a population of roughly 7 million individuals, the data of more than 42 000 individuals would be combined into 21'000 records. This fact should always be kept in mind when exceptional results are found in some statistics.

Moreover, a same individual could have more than one **ID** when a source makes a typing error. Such an error could lead the data collector to define a new individual. Therefore, we recommend before sending data to the data collector, to apply some plausibility tests on the information that define the **ID**.



## 5 A Scheme Providing Weak Anonymity

In 1997, the federal office for statistics proposed a privacy-preserving data collection scheme [1].

It was intended to be used to collect statistical data of in-patient treatments. The scheme is similar to solutions used in France and other European countries. In the following, we summarize the proposed implementation and analyze it with respect to the aforementioned requirements.

### 5.1 An Informal Overview

In this section, we give an informal overview of the proposed scheme. For technical details, we refer to [1].

On a conceptual level, the scheme assumes a simpler system model than the one we presented. In particular, it only considers two types of entities, the sources and the central data collector. The recovery authority as well as data analysts are not considered as explicit entities of the model. Rather, these entities are an implicit part of the data collector.

#### The Data Transfer Protocol:

To define the data transfer protocol, we use a cryptographic one-way function. Such a function has the following properties:

- The input can have any length
- The length of the output is constant (for example, 160 bit)
- For a given input, there is only one output
- For a given output, it is infeasible to compute an input which gives this output

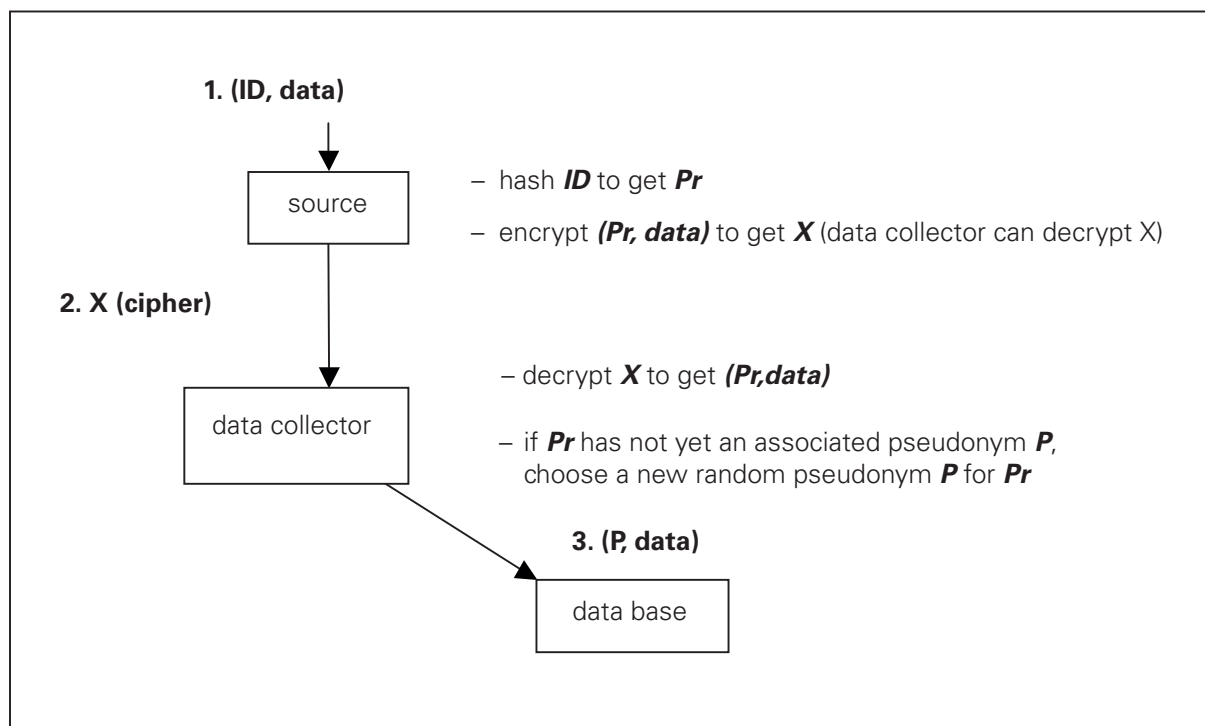
Let  $h(\cdot)$  denote a such a one-way function.  $h(\cdot)$  maps an identity  $ID$  on a random bit-string  $Pr$ .  $Pr$  is here a pseudonym that we cannot distinguish from a random number. Consider a source  $S$  holding a medical record  $(ID, data)$ . The basic idea of the data transfer protocol is as follows (see Figure 3):

1. The source computes a temporary pseudonym  $Pr$  as  $h(ID)$ , and sends  $(Pr, data)$  encrypted to the data collector.
2. Upon receiving this message, the data collector decrypts it and checks if it has already given out a pseudonym  $P$  for  $Pr$ . If not, it chooses the pseudonym  $P$  at random, and remembers the matching  $(Pr, P)$ .
3. The data collector stores the tuple  $(P, data)$  in its database.

Technically, choosing a random pseudonym  $P$  and remembering the matching  $(Pr, P)$  is done by choosing a secret key  $k$  at system initialization, and then computing the pseudonym  $P$  as the symmetric encryption  $sym_k(Pr)$  of  $Pr$ , using for example the IDEA [9] or the AES encryption scheme.

Using this technique, the secret key  $k$  becomes the key to recover identities of a pseudonym, and thus, deserves special protection. To address this issue, the authors suggest to share the key  $k$  among say three individuals, who would have to pool their key-shares whenever the key is needed.

Figure 3 Data Flow in the Data Transfer Protocol



### The Data Export Protocol:

As mentioned above, this scheme does not have the concept of data analysts, but assumes that all analysis is done directly on the database of the data collector. This introduces some security issues, as the analysts have access to **all** collected records, and not only on a subset of the records.

The scheme mentions that it would be possible to give data to external analysts, such as universities, or health care providers. These external analysts are a similar concept as the data analysts in our system model. Yet, to export data, the scheme does not foresee a special protocol, but simply dumps the data directly from the data collector's database.

As a consequence, external analysts that perform regular analyses (say once every year), will gradually get a bigger picture on the individual medical histories. To address this issue, the scheme proposes to refresh the pseudonyms in the database periodically (for example, once a year), so that linkage of records cannot be done across multiple years.

To do this refresh, the data collector computes for every pseudonym  $P$  in the database the corresponding  $Pr$ , chooses a fresh pseudonym  $P'$ , and exchanges in the database all pseudonyms  $P$  with  $P'$ . Furthermore, it remembers the new matching  $(P', Pr)$  for future updates.

Technically, the refresh is done by first decrypting all pseudonyms  $P$  with key  $k$ . This gives all temporary pseudonyms  $Pr$ . Then, the collector chooses a fresh key  $k'$ , and computes the fresh pseudonyms  $P'$  as the symmetric encryption of the  $Pr$ 's under the fresh key  $k'$ .

### The Recovery Algorithm:

The recovery algorithm in this scheme is conducted by the data collector and the source together. To resolve the identity behind a pseudonym  $P$ , the data collector first looks up the corresponding temporary pseudonym  $Pr$ . It then asks a source for a list of all its patients, and checks for every  $ID$  if  $Pr = h(ID)$ . If this holds for some  $ID_0$ , then this is the sought identity.

## 5.2 Verification of Requirements

It is easy to see that the scheme satisfies link-ability, recoverability, and weak anonymity. While this provides already some protection of the privacy of the involved patients, there are still some issues that remain unsolved. Specifically:

Merging the data collector with the recovery authority and the data analysts induces a security / availability problem. On one hand, one has to trust the data collector. Thus, it needs to be highly secured and handled by a restrictive access policy. On the other hand, the data collector has to be accessible for all the data analysis, requiring availability.

This conflict becomes even more apparent if the secret key  $k$  used by the data collector to choose the pseudonyms is protected by a sharing scheme. Specifically, whenever the data transfer protocol is launched, the individuals holding a key share need to meet in person, pool their shares, and make the key available to the data collector. If data transfer protocols happen rarely, say once a year, this may be feasible. Yet, it does not scale to regular data collection (say, once every month, or every week).

### The data collector as big brother:

The scheme provides only weak anonymity (see arguments below), but not strong anonymity, giving the data collector all power to observe individual medical histories. To see this, notice that the data collector can trivially link a given identity  $ID$  to a pseudonym  $P$  by simply computing  $Pr = h(ID)$  and  $P = sym_k(Pr)$ . In other words, if the data collector is interested in the medical career of some person, it can have a look completely on its own.

Similarly, the data collector can deduce the identity from a pseudonym  $P$  on its own, even though some more effort may be involved: From  $P$ , the data collector can derive  $Pr$  and now has to find an  $ID$  such that  $Pr = h(ID)$ . This is easy, as the space of  $ID$  is only as large as the population of Switzerland. Given that there are only about 7 millions people, launching a brute force attack over these  $ID$ 's is trivial. If such a scheme is chosen in practice, one has to define very carefully who will be the administrator of the data collector, since a person who has a full access to the data collector system could play the role of big brother.

### Data analysts:

When data is passed over to external analyst, the pseudonyms  $P$  are not altered. This allows several external analysts, each holding a data subset from the data collector's database, to combine their data and thereby augment their view on the individual histories.

This does not affect the anonymity as long as the pseudonym is not resolved. Yet, it makes it hard to enforce that a certain data analyst may only get a certain subset of data. Refreshing pseudonyms on a yearly basis as proposed by the system addresses this problem, yet may be too coarse grained.

### Integrity:

Integrity is not guaranteed as the data collector accepts messages without verifying any signatures, and thus, would also accept a message that was never sent by a source but by some malicious adversary that has access to the network. We remark that this could be easily fixed by using SSL/TLS connections with client authentication between the sources and the data collector. Yet, in the original scheme, one also intended to allow data being transferred by a physical device (as for example a CD-Rom). In this case, an explicitly authentication code by the source is needed to guarantee authenticity of the data.

## 6 A Scheme Providing Strong Anonymity

In the following, we propose a solution that provides strong anonymity. Our scheme follows closely the lines of the previous scheme, yet overcomes its limitations by the following main ideas:

- In the previous scheme, the data collector not only collects the data, but also takes on the role of the recovery authority and data analysts. We factor these specific tasks out of the data collector, and assign them to the corresponding entities. This is essential to establish strong anonymity.
- The previous scheme exported data directly from the data collector's database. We use a special export protocol, which makes anonymous these records for every export, thereby making it harder for two data analysts to merge their sets and augment their respective views on individual medical histories.
- The previous scheme did not authenticate sources. We base our system on a public key infrastructure that allows us to authenticate all involved entities. This way, our scheme guarantees integrity of the data.

### 6.1 An Informal Overview

The following is a conceptual overview of our scheme that blends out the cryptography behind it. For details on the actual implementation, we refer to the next section.

#### The Data Transfer Protocol:

Consider a source holding a medical record  $(ID, data)$ . The basic idea of our data transfer protocol is as follows (see Figure 4 for an illustration):

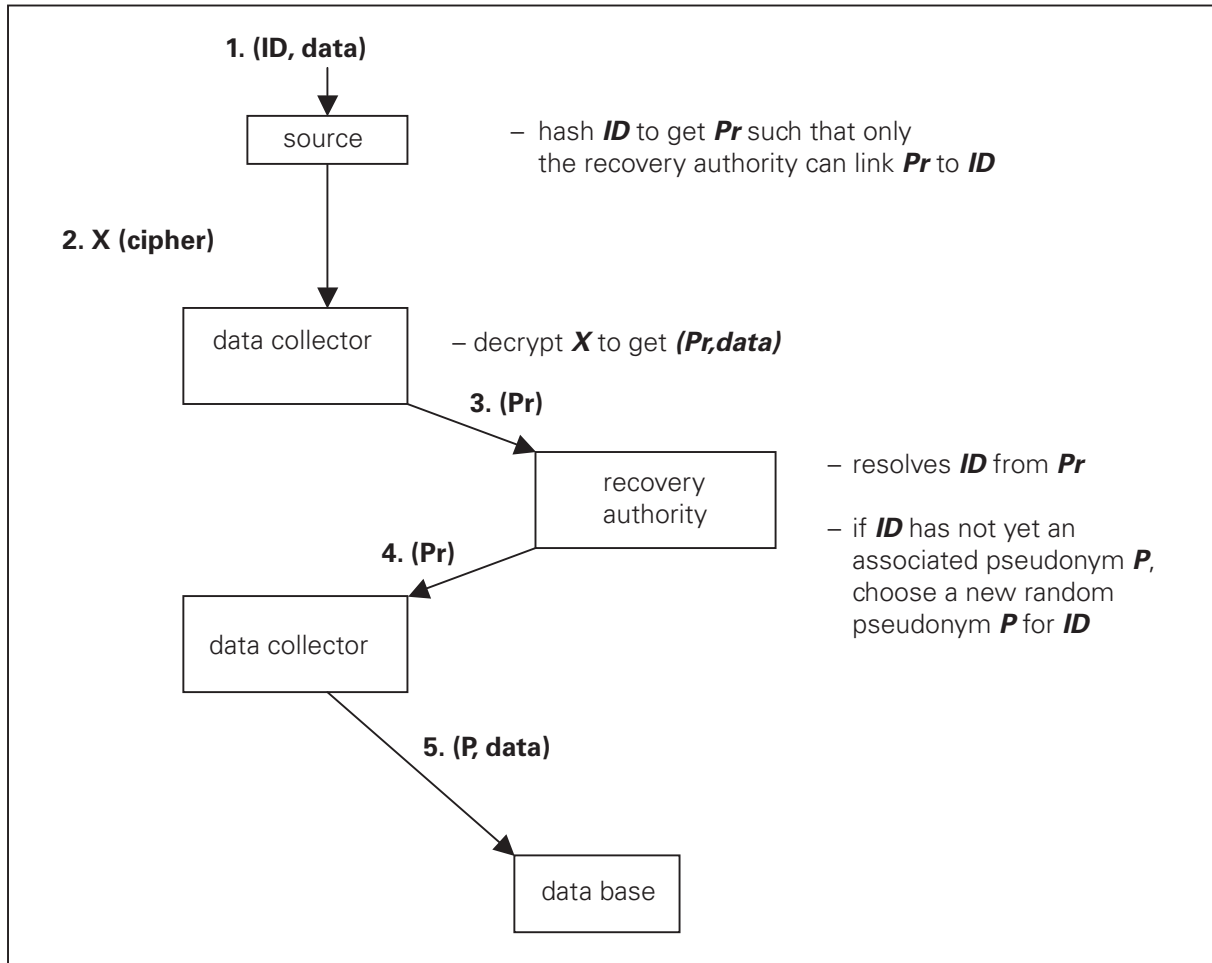
1. The source first generates a pseudonym  $Pr$  such that **only** the recovery authority (and the source itself) can link  $Pr$  to  $ID$ . It then sends  $(Pr, data)$  to the data collector.
2. When the data collector receives the message  $(Pr, data)$ , it asks the recovery authority for a unique pseudonym  $P$  for the patient hidden by the pseudonym  $Pr$ .
3. When the recovery authority receives such a request from the data collector, it derives the patient's identity  $ID$  defined by  $Pr$ . It then checks if it already has given out a pseudonym  $P$  for this patient. If not, it chooses a pseudonym  $P$  at random, remembers the relation  $(Pr, ID)$ , and sends  $(Pr, P)$  to the data collector.
4. When the data collector receives  $(Pr, P)$ , it stores  $(Pr, data)$  in the database.

Since the recovery authority chooses the pseudonym  $P$  for a patient  $ID$  at random, the protocol guarantees that neither the source nor the data collector can link the pseudonym  $P$  to the identity  $ID$  (and vice versa). Furthermore, since the recovery authority chooses only one pseudonym for every identity, the protocol ensures that if two sources transfer a record for the same patient  $ID$ , then the data collector will store these records under the same pseudonym.

Notice that we do not address integrity in this high-level description. One can guarantee integrity by having the sources sign the payload on their data, and appending this signature to the records. See section 6.2 for details.

In practice, it may occur that a record has to be deleted. To do this, the source can use the same data transfer protocol, but mark the record being transmitted with a flag **delete**. The data collector then deletes the record under the corresponding pseudonym, rather than inserting it.

Figure 4 Data Flow in the Data Transfer Protocol

**The Data Export Protocol:**

The basic idea of our data export protocol is as follows. For every pseudonym **P** that occurs in the data records to be exported, the data collector chooses a fresh pseudonym **P'** at random. It then replaces all pseudonyms **P** with their fresh counterpart **P'** in the records to be exported, and sends the records to the data analyst.

Apart from these records, it also sends a token to the data analyst that can later ONLY be used by the recovery authority to resolve the original pseudonym **P** from a fresh pseudonym **P'**. Refreshing the pseudonyms for every export has the advantage that two different data analysts cannot pool their data to augment their historical data on individuals.

**The Recovery Algorithm:**

The recovery algorithm is straight-forward: Suppose the recovery authority receives a pseudonym **P'** and a token from a data analyst. By the definition of this token, the authority derives the original pseudonym **P** as stored by the data collector. Since the recovery authority itself issued this pseudonym, it can easily derive the corresponding identity **ID**.

## 6.2 Technical Details

Our solution heavily relies on the existence of a PKI. Specifically, we assume that every source  $S_i$ , the central data collector  $C$ , the recovery authority  $R$ , and every external data analysts  $A_i$  have a private and public key-pair for encryption and another one for signing. For a user  $i$ , we will denote his two encryption keys by  $ep_i$  (public) and  $es_i$  (private). His signing keys will be denoted by  $sp_i$  (public) and  $ss_i$  (private).

We assume that the encryption scheme enjoys **semantic security (under adaptive chosen message attacks)**. Intuitively, this means that it is impossible to distinguish the encryption of two known messages  $m_1$  and  $m_2$  if the private key is unknown. RSA-OAEP [6] is an example of such a scheme. We denote the encryption of a message  $m$  under a public key  $ep_i$  by  $c = enc(ep_i, m)$ .  $c$  is called the cipher of the message  $m$ . The decryption of a cipher  $c$  will be defined with  $m = dec(sp_i, c)$ .

We assume that the signature scheme is **existentially unforgeable (under adaptive chosen message attacks)**, which means that it is impossible to generate a message and signature pair without knowing the corresponding private key. We denote the signing of a message  $m$  under a private key  $ss_i$  by  $s_m = sig(ss_i, m)$ , and the verification of a signature  $s_m$  on a message will be defined by  $ver(sp_i, s_m, m)$ , which outputs either one for valid, or zero for invalid.

We also assume that all entities are connected to the data collector by secure links (this is easily achievable given the PKI).

### Data Transfer Protocol:

Let  $(ID, data)$  denote the record that a source  $S$  intends to transfer to the data collector  $C$ . The source therefore encrypts  $ID$  under the public key of the recovery authority, and sends this encrypted identity  $Pr$  along with the corresponding record data to the data collector.

Moreover, to insure the confidentiality of  $data$  on the channel between the source and the data collector, one can cipher this information using the public key of the data collector. We don't include this function in Data Transfer Protocol, but we recommend in practice to apply SSL/TLS protocol between all mentioned clients and servers.

The data collector forwards  $Pr$  to the recovery authority, which decrypts  $Pr$  to get  $ID$ . It then encrypts  $ID$  under a symmetric key  $k$ , and sends this encryption  $P$  to the data collector. Upon receiving this message, the data collector then stores a tuple  $(P, data)$  in its database.

We remark that the symmetric key  $k$  used by the recovery authority is the same for every identity  $ID$ , and is chosen at system initialization.

The protocol has to be extended slightly such that the recovery authority only provides pseudonyms for  $Pr$ 's that were produced by a source. Otherwise, the data collector could simply compute  $Pr$  for the patient he is interested in, and ask the recovery authority for the corresponding pseudonym. To prevent this, the source appends a signature to  $ID$  before it encrypts and sends it to the data collector. Furthermore, the recovery authority only provides the pseudonym if the request comes from a valid source.

Algorithm 6.2 summarizes the computational steps in more detail. We write  $sym_k(m)$  to denote a symmetric encryption of a message  $m$  under a key  $k$ .

**Algorithm 1** The Data Transfer Protocol**source S:**

- input ( $ID$ ,  $data$ )
- compute  $s_{ID} = \text{sig}(ss_S, ID)$  and  $s_{data} = \text{sig}(ss_S, data)$
- compute  $Pr = \text{enc}(ep_R, ID || s_{ID})$ , where  $||$  denotes concatenation
- send the message  $(Pr, data, s_{data})$  to the data collector  $C$

**data collector C:**

- receive message  $(Pr, data, s_{data})$  from source  $S$ :
- if  $\text{ver}(sp_S, s_{data}, data) = 0$  then abort
- send  $(Pr)$  to the recovery authority  $R$
- wait for receiving an answer  $(Pr, P)$  from  $R$
- if  $P = \text{NULL}$  then discard the record
- else store  $(P, data)$  in the database

**recovery authority R:**

- **initialization:** choose a random secret key  $k$  for the symmetric encryption algorithm  $\text{sym}$
- receive  $(Pr)$  from the data collector  $C$ :
- compute  $(ID || s_{ID}) \leftarrow \text{dec}(es_R, Pr)$
- if  $\text{ver}(sp_S, s_{ID}, ID) = 0$  then send  $(Pr, \text{NULL})$  to  $C$
- else compute  $P \leftarrow \text{sym}_k(ID)$
- send  $(Pr, P)$  to  $C$

**Remarks:**

The above protocol would put quite some load on the recovery authority if it is invoked for every data record separately. One can avoid this by processing the records in batches, which contain all records of a certain time period (say one day, or one week). In such a batch, there may be many records for the same patient  $ID$ . If this is the case, the source may produce only a single encryption  $Pr$  per  $ID$  per batch, and re-use  $Pr$  for every record of patient  $ID$ . This way, the data collector would have to query the recovery authority only once for each patient in a source-batch.

We could also allow the source to send  $(Pr, data)$  directly to the recovery authority, which then follows the protocol as described. However, we do not recommend this as it would put a high load on the recovery authority, as all the payload data goes through it, and the above batching would not be possible anymore. Furthermore, to ensure the security of the recovery authority, it makes sense to restrict access to this entity to only the data collector.

### Data Export Protocol:

Suppose a data collector **C** needs to export a set of records to a data analyst **A**. If the data analyst requests unlinked-data, i.e., data that does not allow to link records of the same individual to each other, then **C** sends all records requested directly to **A** with the pseudonyms **P** replaced by **NULL**.

If the data analyst requests linked-data, then **C** first chooses a random symmetric encryption key **k**. It then sends for every record  $(P, data)$  a record  $(sym_k(P), data)$  to **A**. After the last record, it encrypts the key **k** under the public key of the recovery authority **R**, and sends this encrypted key  $k' = enc(ep_R, k)$  on a token to the data analyst. If necessary, the token will allow later to recover the identity behind a specific pseudonym with the help of the recovery authority.

### Recovery Algorithm:

The recovery algorithm works as follows. The recovery authority **R** takes as input  $k'$  on the token and a pseudonym  $P' = sym_k(P)$  from a data analyst. It then computes

$$k \leftarrow dec(es_R, k')$$

and outputs the patient's identity as

$$ID \leftarrow dec(es_R, sym_k^{-1}(P'))$$

### A Remark on Supporting Multiple Identities:

A data record may contain not only the patient identity, but also the identity of the treating doctor, or of the treating nurse, and so on. It may be desirable to make anonymous these identities, as well. This can be done using the same techniques as we use to make anonymous the patient's identity.

Care has to be taken, however, that a person, which is both, a doctor and a patient, receives a different pseudonym for each of its roles. This can be guaranteed as follows. Recall that the recovery authority computes the pseudonym for an identity **ID** by encrypting it under a key **k** using a symmetric encryption scheme.

If we have multiple roles of identities  $ID_1, \dots, ID_n$  that have to be anonymous, then the recovery authority has to use different keys  $k_1, \dots, k_n$  to generate the pseudonyms, one key for each role.

## 6.3 Verification of Requirements

We have to show that link-ability, strong anonymity, recoverability, and integrity holds as defined in Section 3.2.

### Link-ability:

Consider two records  $(ID, data)$  and  $(ID', data')$  that two sources send to the data collector through the data transfer protocol. We have to show that if  $ID' = ID$ , then the data collector eventually stores two records  $(P, data)$  and  $(P', data')$  with  $P = P'$ .

When the data collector receives  $Pr = enc(ep_R, ID || s_{ID})$  and  $Pr' = enc(ep_R, ID' || s_{ID'})$ , it forwards both pseudonyms **Pr** and **Pr'** to the recovery authority **R**. Using the encryption scheme, **R** will decrypt both pseudonyms to the identity **ID**, and thus, will send the same pseudonym **P** to the data collector.



**Strong anonymity:**

Consider a special scenario, where we only have one source treating only two patients  $ID$  and  $ID'$ . Suppose the two patients are mapped to pseudonyms  $P$  and  $P'$  by the recovery authority, and suppose the data collector knows the patient identities, but not the mapping to their pseudonyms. We have to show that if the source follows the protocol, and if neither of the pseudonyms is resolved by the recovery authority, then the data collector and the data analysts together can guess the correct identity with probability  $1/2$ .

Since the data analysts get all their data from the data collector, it suffices to show that the data collector on its own cannot guess the identity behind the pseudonyms with more than probability  $1/2$ .

We show this by contradiction, i.e., we show that if the above does not hold, we could use the data collector to break the semantic secure encryption scheme used to encrypt the identities. Recall that semantic security means that it is infeasible to distinguish the two random variables  $(m_1, m_2, c_1, ep_k)$  and  $(m_1, m_2, c_2, ep_k)$ , where  $c_1$  and  $c_2$  are the encryptions of  $m_1$  and  $m_2$ , respectively, under the public key  $ep_k$ .

Given a successful data collector (i.e., a collector that can link the pseudonyms to the identities), we build a distinguisher for the semantic encryption scheme as follows. We use the data collection with the given data collector as a sub-component as follows. We first define  $ID$  and  $ID'$  to be equal to  $m_1$  and  $m_2$ . By hypothesis,  $m_1$  and  $m_2$  are known. Let the public key of the recovery authority be  $ep_R$ . We then feed the source with two records  $(ID, data)$ , and  $(ID', data')$ , and let the system run. Notice that we do not know the private key  $es_R$  belonging to the recovery authority. Thus, the data analyst cannot derive the identities by decrypting the pseudonyms  $P$  and  $P'$ . However, since we control the entire system (except for the data collector), we know which pseudonym belongs to which identity without decrypting. So we can simply let the recovery authority choose the pseudonyms for each patient at random and send them to the data collector.

So far, the data collector cannot tell a difference between this 'simulated' environment and the real environment, as the distribution of all the messages (in particular, the pseudonyms  $P$  and  $P'$ ) are identical. This does not change, either, if the data collector tries to query the recovery authority on any self-computed encryption of an identity  $ID$ , since by the security of the signature scheme used by the source, the data collector cannot compute a signature on  $ID$  which would be required by the protocol.

This implies, together with the assumption that  $C$  is successful in the real environment, it will also be able to predict which pseudonym corresponds to which patient with more than probability  $1/2$ , i.e., it can distinguish  $(ID, ID'; P, ep_R)$  from  $(ID, ID'; P', ep_R)$  with non-significant probability. We can use this output directly to break the semantic security of the encryption scheme.

A similar argument can be given to show that also the source cannot link a pseudonym  $P$  to one of its patients  $ID$ .

**Recoverability:**

Recoverability follows directly by the protocol.

**Integrity:**

We have to show that if the data collector receives a record  $(P, data)$  from the data transfer protocol, then the recovery authority maps  $P$  to some identity  $ID$  such that there exists a source which provided the record  $(ID, data)$  as input to the protocol.

We start with the fact that the data collector received a record  $(P, data)$ . By the protocol, it must have asked the recovery authority for a pseudonym given some  $Pr$ . Thus, there exists some  $ID$  such that  $P = sym_k(ID)$ , where  $k$  is the key used by the recovery authority. This implies the first part of integrity.

To see the second part, note that if the data collector receives as output a record  $(P, data)$  from the data transfer protocol, we may assume that it received a record  $(Pr, data, s_{data})$  from some source  $S$ , where  $s_{data}$  is a valid signature on  $data$  under the public key  $sp_S$ . It follows by the unforgeability of the signature scheme that source  $S$  has transmitted the correct record as input to the protocol.

## 6.4 Offline Recovery Authority

Notice that in the proposed scheme, the recovery authority has to be online during the data transfer protocol. On the other hand, it needs to be very secure. Since these two requirements conflict to some extent, it would be desirable to have a data transfer protocol where the recovery authority does not have to be involved and can stay offline. In such a system, the authority could live in a safe-like room, completely disconnected from any network.

Despite these advantages, an offline scheme also has some security drawbacks: If some source and the central data collector behave malicious, they receive essentially the power of the recovery authority. This is because they can compute — by definition of an offline scheme - the pseudonym of an arbitrary patient without interacting with any other entities. Since there are only a limited (and well-known) set of patients, they can easily compute a complete mapping between identities and pseudonyms.

We remark that theoretically, this is also possible in any online scheme. This is because one cannot verify if a data record provided by a source represents a treatment that indeed happened at that source. Thus, a malicious source can simply invent a data record for every patient, insert it into the system, and compare it with the output of the data collector to build a complete mapping between identities and pseudonyms (see Section 3.3 for details).

In this attack, however, the recovery authority is involved for every mapping that is computed. Thus, in practice it should be feasible to detect such an attack by analyzing the logs on the recovery authority side, except if many sources are malicious. Furthermore, a medical card for each patient could allow the introduction of a new feature in the protocol: one can ask a confirmation of the patient about the integrity of a data record, and thus prevent such attacks completely.

Notice that this drawback heavily relies on the fact that one entity receives exactly one pseudonym. If this link-ability is not required, the above arguments do not hold anymore, and an offline scheme is certainly the desired approach. Yet, if link-ability is required, we recommend to use an online scheme as the one proposed in the previous section.

## 7 The Strong Scheme in Practice

In this section we present a possible implementation of the scheme providing strong anonymity (section 6). This implementation has several requirements concerning system architecture and performance. First, we give an overview over the services, their interactions and locations. Then we present an estimation of hardware requirements for these services.

### 7.1 System Architecture

Three main areas have to be distinguished:

- A database area, where the main database is located together with access controlling systems.
- A recovery area, which must be located outside the database area for security reasons.
- The Internet, from where requests for access to the database area origin.

Sources connect over the Internet to the database area to submit their data. They have to pass a reverse-proxy where all requests must be authorized and authenticated before they are passed on. Data-analysts using an export client can only query the database by connecting over the reverse-proxy, where they are also authenticated. This allows to build only one entry-point to the database area and have all other connections in its own network, providing higher security than having several entry-points.

The proxy server passes all requests either to a server dedicated to import data from sources or to a server dedicated to query the database. These servers have only the minimal needed functionality for their job: The import server has only write-access to the database (except read-access on some log-tables to ensure no data is committed twice), while the export server has only read-access. This modularity allows to extend the performance of the system by adding additional export or import servers and having the reverse-proxy act as a load-balancer.

The import server must be able to communicate with the recovery to receive the pseudonyms **P**s. We propose to connect the import server directly with the recovery over a dedicated leased line, protected on both side with a firewall.

**Figure 5** gives an overview over the proposed servers. All data transfers should be handled over **SSL/TLS** connections with server and client authentication. If the load on the import and export servers are low, they could be merged into one machine.

The data has not only to be secured to unauthorized access, but also to erroneous manipulation and hardware failure. We propose thus to use database-replication tools to have always an updated copy available. Additionally, it would be useful to create backups regularly to a removable device for creating remotely storable backups. This could be done using tape decks or a similar technology. The records should be further protected by using an encrypting hard disk or file system for the database storage.

The database can get quite large. To store the data, a **RAID-System** (Redundant Array of Inexpensive (or Independent) Disks) seems inevitable. We propose to use a **RAID 0+1** or a **RAID 5** system to create a large and failure-tolerant disk-set.

Note that this concept does not ensure 100 % up-time, but it protects quite effective from data loss. Because the system is mainly used as a data collecting service and (indirectly) for data analysis, it is not critical if the service is not available for a day or two. But it would be much worse if the data of the last five years would be lost due to a hard-disk crash or a similar incident.

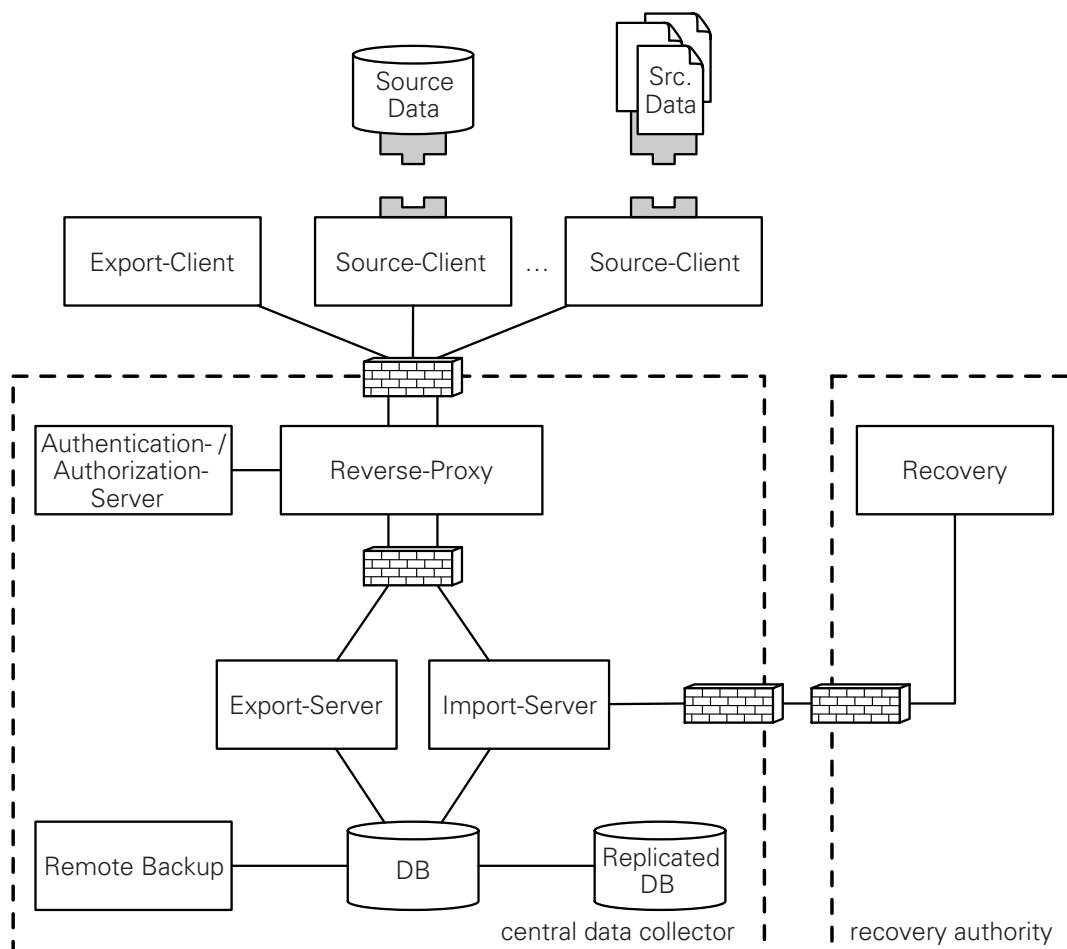
The different sources use many different internal systems to manage their data. It is far beyond the scope of this project to unify all those systems. Thus the sources must be able to generate correctly formatted records from their systems on their own to submit them to the central data collector. For security reasons and reasons of manageability, the sources could receive a client software which implements the protocol to submit the records to the data collector. The client offers one or more ways to load the records, e.g. reading flat files (**CSV** or similar), selecting the data from a database-table over **ODBC**, or a similar standards-based interface. This client application marks the outermost part of the system that tangents the sources.

## 7.2 Performance Estimations

Collecting the data from all hospitals, health insurance funds and care providers generates quite an amount of data. Assuming we have every year 5 million patients, each being treated for 6 diseases. Every treatment of a disease generates 100 records in average. This means that every year 3 billion records have to be transmitted to and stored in the database. To simplify the following calculations and to allow for a potential future increase in records, we will use the number of 5 billion records per year (about 13.7 million records per day). This already shows that the data has to be transferred regularly and not only once a year as in [1].

We assume that an average record has a size of 1 KB. This may sound huge in the first moment, because in every record only a few values are submitted. But as soon as one or more values are encrypted in the records, they require much more place. Storing all those records requires thus about 5 TB for each year (about 14 GB for each day). This can be easily handled by today's **RAID**-systems or networked storage.

Figure 5 Proposed Hardware Strategy



All these records have to be transferred over the Internet. The 14 GB per day correspond to a steady flow of about 160 KB per second, or 1.3 Mbit/s. Assuming a peak of 15 to 20 times the throughput, the data collector must be connected using a 20 Mbit lane. Assuming that about 100 sources would deliver the data, the load for each could easily be handled by the upstream of today's broadband connections.

As the aforementioned numbers show, the physical transmission is not a problem. But before and after transmission, in every record some data must be sign and encrypted, or decrypted and the signature verified. Table 1 shows for each entity how many cryptographic functions must be applied per record, while table 2 shows the number of cryptographic functions applied for each entity per day.

We can then calculate the time a server is busy to perform the cryptographic functions by making an assumption about the number of functions a modern server can perform today (see table 3).

The times are shown on table 4. The final daily loads are summed up in table 5. As can be seen, the recovery is the most heavily used component. Note that these numbers are only rough estimations, and that no hardware acceleration or optimizations were considered. A possible optimization for the recovery is to cache the calculated pseudonyms. Processing over 13 million records on a day while there are only 5 million patients, the load can be cut at least in half in theory. In practice, the factor will be even larger because it is unlikely that all 5 million patients are treated in the same day or same week.

Table 1 Number of cryptographic functions applied per data record

	sign	verify signature	asymmetric encryption	asymmetric decryption	symmetric encryption/ decryption	SSL/TLS (1 KB-Block)
source	1		1			1
collector		0.001				3
database						1
recovery		1		1	1	2

Table 2 Number of cryptographic functions applied per day

	sign	verify signature	asymmetric encryption	asymmetric decryption	symmetric encryption/ decryption	SSL/TLS (1 KB-Block)
source	137k		137k			137k
collector		13.7k				31.1m
database						13.7m
recovery		13.7m		13.7m	13.7m	27.4m

Table 3 Assumed performance of a modern single CPU server (2005)

	sign	verify signature	asymmetric encryption	asymmetric decryption	symmetric encryption/ decryption	SSL/TLS (1 KB-Block)
number of functions per sec.	500	10'000	10'000	500	250'000	20'000

Table 4 Assumed time in seconds to handle daily load for each function

	sign	verify signature	asymmetric encryption	asymmetric decryption	symmetric encryption/ decryption	SSL/TLS (1 KB-Block)
source	274		14			7
collector		1.3				1'555
database						69
recovery		14'000		27'400	55	1'370

Table 5 Average Processor Loads

	total time in seconds	load over 24 hours
source	295	0.34 %
collector	1'556	1.8 %
database	69	0.08 %
recovery	41'285	48 %

## 8 Further Applications and Extensions

The privacy-preserving data collection offers interesting applications for example for payment, controlling and risk balancing. In combination with a medical insurance card even anonymous treatment or access to an online medical data collections would be possible.

### 8.1 Medical Insurance Card

On the 22nd of June 2005, the federal council passed a concept for the initiation of a swiss medical insurance card. This concept comprises that from 2008 every insurant has to present his card for drawing any medical benefits. We highly recommend to provide such medical insurance cards with a chip having cryptographic capabilities (as for example a smart card [4, 7]). This would open many very interesting possibilities.

- **Unique Identification:** A medical card is the ideal medium for saving a unique identifier as described in section 4. It can even be used to generate, per treatment, a unique pseudonym which will be used for the privacy-preserving data collection.
- **Anonymous Treatment:** A medical insurance card does not need any imprints of personal data. So, a patient could take his card and go to a care provider and ask for a treatment. Based on a zero-knowledge proof [8], the care provider verifies that the patient has a valid insurance. After verification, the care provider only receives pseudonym which was automatically created by the card. After the treatment, the care provider sends the bill to a central billing agency. This agency debits the costs to the accurate insurance and forwards the detailed data to the privacy-preserving data collection.
- **Online Medical Data:** A medical card could be used for authorization granting dedicated access to an online database containing detailed medical data about its owner as for example x-ray pictures, blood group, and prescribed medicaments.

### 8.2 Payment and Controlling

Payment of care providers is still based on two old models.

#### 1. Tier Payant:

The care provider directly sends the bill to the accurate insurance. The patient never sees a bill and has no possibility of control.

#### 2. Tier Garant:

The care provider sends the bill to the patient. The patient pays the bill and sends it to the insurance for refunding.

The privacy-preserving data collection could be used as central billing agency.

### 8.3 Balance of Risk

A very hot topic not yet mentioned is the balance of risk between the insurance carriers. Right now there is an ongoing political and juristic discussion on how to justly revise this balance. In this discussion, the privacy-preserving data collection offers very interesting possibilities. Its detailed data can be used for risk calculation and balancing between the insurance carriers.

## 9 Discussion and Conclusion

In this work, we investigated the problem of a privacy-preserving data collection, and proposed a system that offers a reasonable protection for the privacy of the individuals whose data is collected.

A distinguishing property of our system is that it has a designated entity, the recovery authority, which is responsible for the anonymity of the individuals. Specifically, even the central data collector cannot recover the identity behind a pseudonym.

This contrasts with the previous system proposed in [1], where anonymity was established and guaranteed by the central data collector, and no recovery authority existed.

We believe that having to trust the central data collector is a major drawback of this system. This is because the data collector is likely to be a complex system, given the number of clients it communicates with, and the amount of data it collects. In practice, such a system is harder to secure than a recovery authority, which has a very specific task and can be designed in a simple and effective way.

Another difference to the previous system is that we factor data analysis out of the data collector, i.e., rather than allowing data analysis directly on the data collector's data, we suggest to use an export protocol which re-randomizes the data collector's pseudonyms before the data is given to an analyst.

This ensures that two data analysts cannot pool their data and thereby augment their view on the individuals, as the same individual will have different un-linkable pseudonyms for every data analyst.

And finally, we point out the importance of a mechanism to verify that a certain data record being collected is genuine, i.e., that the record represents an event that actually happened in the real world. Specifically, without such a mechanism, anonymity cannot be ensured with respect to malicious data providers (sources).

At the end of this report, let us quickly glance at Denmark. They already have successfully implemented an overall medical database (Medcom) [3] in the late eighties. Medcom offers open access (the patients are not anonymous!) on the whole data for care providers and statisticians. Such a model is unthinkable for Switzerland with its strict data protection law and its background, especially due to the 'Secret Files Affair'. But the example of Denmark should encourage us to realize a privacy-preserving data collection which fulfills all necessary requirements.



## References

- [1] Bundesamt für Statistik. Der Datenschutz in der Medizinischen Statistik - Statistik der stationären Betriebe des Gesundheitswesens, 1997.
- [2] Bundesamt für Statistik. Gesundheitsversorgung, Kosten, Finanzierung - Kennzahlen, 2005.
- [3] Markus Hofmann. NZZ Folio: Spitzenposition für Danemark, September 2005.
- [4] International Telecommunication Union (ITU). Public-key infrastructure (x.509).
- [5] David-Oliver Jaquet-Chiffelle. Description des méthodes envisagées pour protéger la confidentialité des données personnelles. 1997.
- [6] RSA Laboratories. PKCS #1: RSA Cryptography Standard, 2003.
- [7] RSA Laboratories. Public-key cryptography standards #11 and #15, 2003.
- [8] S. Micali S. Goldwasser and C. Racko. The knowledge complexity of interactive proofs systems. Proceedings of the 17th Annual ACM Symposium on Theory of Computing, pages 291–304, 1985.
- [9] J. L. Massey X. Lai. A proposal for a new block encryption standard. pages 389–404. Springer-Verlag, 1991.

