

All reproduction is prohibited but direct link to the document is accepted:

<http://www.irdes.fr/english/issues-in-health-economics/196-reconciling-general-medical-records-and-health-insurance-reimbursement-data.pdf>

Reconciling General Medical Records and Health Insurance Reimbursement Data: Feasibility Study and First Results

Yann Bourgueil^{ab}, Julie Perlberg^{ab},
in collaboration with Caroline Allonier^a, Philippe Boissault^b, Fabien Daniel^a,
Philippe Le Fur^{ab}, Philippe Szidon^b

^a Irdes ^b Prospere

The aim of this study was to test the feasibility of chaining clinical data with National Health Insurance reimbursement data and to validate the interest in doing so. This data linking trial is part of a long term project aimed at creating an information system enabling the advancement of research on health care services. Based on a representative sample of general practitioners (GPs) and patients, an information system of this kind would provide a means of measuring morbidity management in the primary care sector and a tool to analyse GP practices and patient care pathways so as to improve the efficacy and efficiency of the health care system.

The first step consisted in auditing the technical feasibility of the data linking process. The second phase made it possible to evaluate the interest in enriching Health Insurance data with clinical data in the aim of identifying populations suffering from chronic diseases, namely diabetic and hypertensive patients.

According to ESPS 2010, 85% of respondents (outside institutions) reported consulting a general practitioner at least once during the last twelve months. In 2010, the general health insurance scheme reimbursed almost 234 million general practice consultations, equivalent to 3.7 consultations per inhabitant per year (*Eco-Santé*). This indicates that in terms of numbers, the pri-

mary care sector, and more generally the ambulatory care sector, covers the highest demand for care. The hospital sector covers the proportion of care needs requiring a higher concentration of expertise and technical facilities (one out of five inhabitants hospitalised in 2010, *cf. Eco-Santé*). However, if knowledge of hospital sector morbidity management has progressed over the last twenty years with the creation

of the Information System Medicalization Programme (PMSI, *Programme de médicalisation des systèmes d'information*), the same cannot be said for the ambulatory sector. Apart from the Observatory of General Medicine (OMG, *Observatoire de la médecine générale*) experiment set up by the French Society of General Medicine (SFMG, *Société française de médecine générale*), certain ad hoc

studies such as ECOGEN (Elements of general medical practice consultations conducted by the National College of General Practitioner Teachers research (CNGE, *Collège national des généralistes enseignant*), consultation data collections focused on flu epidemics (www.grog.org/) or studies organised by private companies interested in prescription practices, we have little knowledge of the morbidities diagnosed and treated in private general practice. Yet the clinical situations encountered by GPs are becoming more and more complex. Within the context of an ageing population, multiple concomitant pathologies and loss of autonomy are on the increase rendering treatment more complex (HCAAM, *Haut conseil pour l'avenir de l'Assurance maladie*, 2011). Knowledge of patient characteristics and the diseases recognised¹ and treated in general medical practice are thus necessary conditions for the production of knowledge on effective practices and the adaptation of primary care organisations to achieve more efficient patient care pathways.

Data sources enabling the improvement of knowledge on morbidity in private medical practice

France is one of the rare countries to have developed a comprehensive and centralised information system containing individual medico-administrative and social data managed by public organisms. If the use of these data, essentially produced for management purposes, remain difficult for legal, technical and organisational reasons (Golberg *et al.*, 2012), the public authorities and health system players are seeking to improve data exploitation possibilities in order to increase knowledge and facilitate efficacy and efficiency assessments of the health care system.

Numerous studies, essentially conducted by the National Health Insurance, aim to match beneficiaries' pathologies with reimbursement data using algorithms con-

structed on the basis of knowledge on: reimbursed medication, medical causes leading to registration under the long-term illness scheme (LTI), and hospital diagnoses (CNAMTS, *Caisse nationale d'assurance maladie*, 2014). These methods make it possible to determine costs per pathology and to orient risk management actions.

An alternative approach explored in this study, consists in matching GP consultations and visits with medical problems (symptoms, syndromes and diagnoses) identified by GPs and the resulting prescriptions. Data bases representative of general medical practices such as these have been developed in several countries such as the United Kingdom, New Zealand and the Netherlands (Chevreul *et al.*, 2006).

In the French context, the enrichment of medico-administrative reimbursement data with office-based "medical" data should, in theory, offer new perspectives in the measurement of morbidity encountered in general medical practices and the analysis of health care system performance. It essentially concerns the treatment of multiple concomitant pathologies but also the analysis and understanding of general medical practices and patient care pathways. What percentage of non-insulin dependent diabetics have been identified and treated without medication? What



Conducted within the framework of the multidisciplinary project PROSPERE (Multidisciplinary Research Partnership on the Organisation of Primary Care Services, *Partenariat pluridisciplinaire de recherche sur l'organisation des soins de premiers recours*), this study contributes to the aim of producing knowledge on the performance of primary care services in order to accompany the transformation of the ambulatory care sector*. This project, selected as an emerging team within the framework of the IRESP call for tender in 2008, brings together researchers from IRDES, the French Society of General Practitioners (SFMG) and the Research Centre for Medicine, Science, Health and Society (CERMES). It was financed for a four year period (2009 to 2013) by grant from the CNAMTS and each of the partner institutions.

* In addition to this synthesis, the study will also be published as an article in *Santé publique* (n° 3/2014).

preventive actions have been set up for patients hospitalised for a stroke? How many health problems have been identified and treated in office-based practices among patients hospitalised for decompensated heart failure (ADHF)? What percentage of patients' medication is prescribed and reimbursed following a GP consultation? These and other similar

SOURCES ET METHODS

Two sources

The data used were provided by the Observatory of General Medicine (OMG, *Observatoire de la médecine générale*) and the National Health Insurance Inter-regime Information System (SNIIRAM, *Système national d'information inter-régimes de l'Assurance maladie*). Created by the French Society of General Medicine (SFMG, *Société française de médecine générale*), the OMG, a non-commercial database, is continuously informed in real time as a computerised medical record. The data collected concern all patients having consulted one of the 30 GPs selected for the data linkage experiment at least once during the course of 2008. The data used for the test phase were structured on two levels providing information on patient characteristics on the one hand, and characteristics of the session on the other (consultation or visit). The data extracted from the SNIIRAM database, grouping together all data on health expenditure reimbursements for all patients having consulted one of the 30 selected GPs (outside hospitalisation data that were unavailable in 2008). Only general health insurance regime beneficiaries were able to be identified in the SNIIRAM database in 2008.

Method

Probabilistic data linkage (Silveira and Artmann, 2009) is not based on the unique identification number (NIR) but attempts to link sessions using 6 discriminatory variables: the GPs "Adéli" number, the date and types of session (consultation or visits), the months and years of birth and beneficiaries' gender. The collation of information on dates and types of contact and patient identities allowed us to reconcile the two databases, OMG and SNIIRAM. The data linkage/matching procedure were carried out on the basis of sessions. When a patient's session was matched (identified in both the OMG and SNIIRAM data bases), that patient was considered as matched and all sessions were then integrated into the new paired data set, whether the sessions matched or not. This test was subject to authorisation from the French Data Protection Authority (CNIL, *Commission nationale informatique et libertés*) and a favourable opinion from the Institute of Health Data (IDS, *Institut des données de santé*).

¹ Morbidity is said to be recognised when it is recognised by the health care system to which patients have access.

questions can be explored by linking clinical data and GP practices with patients' reimbursement data routinely produced by the National Health Insurance Inter-regime Information System (SNIIRAM, *Système national d'information inter-régimes de l'Assurance maladie*).

An OMG and SNIIRAM data linkage feasibility test

The project presented here was carried out within the framework of research conducted by the PROSPERE team (Context insert). Its aim was to test the feasibility and utility of creating an ambulatory care research tool at individual beneficiary level by linking medical data provided by an existing general practice data base (OMG) with National Health Insurance reimbursement data (SNIIRAM) using a data linkage method that is not based on beneficiaries' unique identification number (NIR) [Sources and Methods insert].

The first data-pairing phase using the two data bases was carried out on data concerning all patients having consulted one of the 30 GPs selected among the 150 GPs routinely coding for the OMG data base in 2008. This first test evaluated the technical feasibility of this data linkage method.

In a second phase, an exploratory analysis of linked data allowed us to assess the pertinence of chaining these data in the identification of two types of population suffering from chronic diseases, diabetic and hypertensive patients. The clinical information collected during consultations effectively allows the identification of diseases treated at an early stage of their development by GPs, before treatment with medications, diseases recognised as long-term illnesses or hospitalisations related to the disease in question. Quantifying these populations makes it possible to present a more realistic overview of morbidity as encountered by health professionals and thus support them in their preventive actions that are not necessarily based on medication.

In the first phase, the data base created by the linkage of clinical data and SNIIRAM data for the year 2008 is compared with the Permanent Sample of Health Insurance

Beneficiaries (EPAS, *Echantillon permanent des assurés sociaux*) for the same year.

Data linkage for 80% of atients

The OMG and SNIIRAM data bases constituted for the data linkage process counted 37,992 patients and 126,793 sessions in the OMG base, and 35,730 patients representing 117,509 sessions (consultations and visits) in the SNIIRAM database. In total, almost 80% of patients were successfully data matched, or in other words 89,211 sessions for 29,088 patients. Two data bases were then constituted, the first using the session as the base unit, and the other the patient. The data linkage process was carried out on the basis of sessions rather than patients (Sources and Methods). The database linked at "session" level thus only contained matched sessions; that is to say data found in both OMG and SNIIRAM databases. On the other hand, the database matched at 'patient' level contained matched data and data exclusively taken from either the OGM or SNIIRAM databases (Graph 1).

Certain patients present in the SNIIRAM database could not be matched with OMG session data ($n = 6\,642$) (Graph 1). These patients were reimbursed for medical procedures carried out in 2008 by one of the 30 GPs selected, but were not found in the OMG database. This result is probably related to consultation coding problems in the OMG database (coding errors or omissions). As an example, on the totality of sessions, 2,608 (3%) represented visits whereas the actual visit rate was close to 7% of the total sessions recoded in the SNIIRAM base. This gap can probably be explained by a less precise coding of visits compared with consultations in the OMG database.

Inversely, certain patients recorded in the OMG database, that is to say having consulted one of the 30 GPs selected for the OMG during the course of 2008, were not matched ($n = 8\,904$). None of their sessions were recorded in the SNIIRAM database (Graph 1). This can be explained by the fact that patients covered by other regimes than the general health insurance regime were not included in the SNIIRAM data-

base in 2008 [Agricultural Mutual Fund (MSA, *Mutuelle sociale agricole*), Social Security Fund for Self-Employed Workers (RSI, *Régime social des indépendants*), Local Mutualist Sections (SLM, *Sections locales mutualistes*)]. This test phase limitation will be eliminated on subsequent SNIIRAM data exploitation as all regimes of the National Health insurance were taken into account from 2009.

Matched patients comparable with the basic health insurance scheme population in 2008, except in terms of long-term illness scheme beneficiaries

In 2008, data-linked patients were on average 38 years old against 39 years old in the General Sample of Health Insurance Beneficiaries² (EPAS) and in the majority women (54%, proportion identical to EPAS). 2% had benefitted from Universal Health Insurance coverage (CMU, *Couverture maladie universelle*) at least once during the course of the year against 3% in the general population, and 6% from the complementary CMU at least once during the course of the year against 6.6% in the general population (EPAS, 2008). Within the sample, 3,718 (13%) patients were registered as having a long-term illness against 16.5% in EPAS (Païta and Weill, 2008). On average, patients had six contacts with a GP (consultations and visits) during the course of the year, which is close to patients' average use rate in 2008 for general medicine (5.4 contacts on average in EPAS).

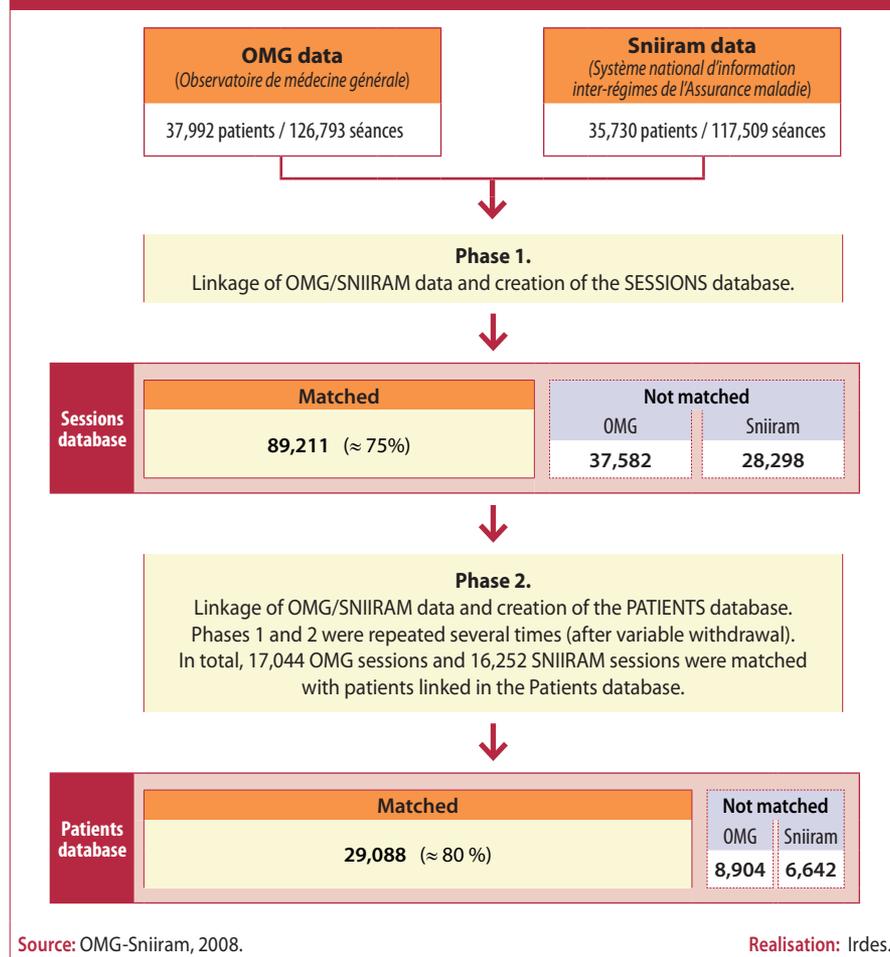
Improved identification of chronic diseases after data linkage

The usability and pertinence of linked data was analysed using two target populations, diabetic patients and hypertensive patients. The populations identified from the LTI "diabetes" or "hypertension" code, or through reimbursed medication specif-

² According to Epas, 69,473 individuals affiliated to the general health insurance regime consulted a GP in 2008.

G1

Description of the OMG and SNIIRAM data linkage process



ic to these pathologies in the SNIIRAM database, were compared with the populations identified through consultation results³ for the same pathologies in the OMG database. In 2008, the year of the test, no information concerning hospitalisations was accessible as the PMSI was not yet linked with the SNIIRAM. This linkage is now operational for all data exploitation after 2008.

The cost of diabetes, one of the major chronic diseases in France, has increased by one billion euros per year over the last few years (Morel *et al.*, 2012). The prevalence of pharmacologically treated diabetes was 4.2% in 2009, that is to say over 2.5 million patients treated in France (Ricci *et al.*, 2010). Diagnosed but

non-medically treated diabetes recorded a 0.6% prevalence rate within the French population in 2009; that is to say approximately 360,000 individuals (Ricci *et al.*, 2010). This population is nevertheless difficult to identify from SNIIRAM reimbursement data only. Furthermore, in 2007 the pharmacological treatment of diabetes represented 9% of total Health Insurance expenditures (Morel *et al.*, 2012). Identifying diabetic patients as early as possible, preferably before treatment becomes necessary, would provide an adapted secondary preventive action and delay progression of the disease and the need for medication.

Using the paired data set created for this study, the comparison of patients registered under the long-term illness scheme (LTI) with consultation results showed that 94% (521/553) of LTI "diabetes" patients were specifically treated for their diabetes in 2008 by one of the 30 selected GPs (Sources and Method). In contrast, only 65% (521/799) [Graph 2] of patients

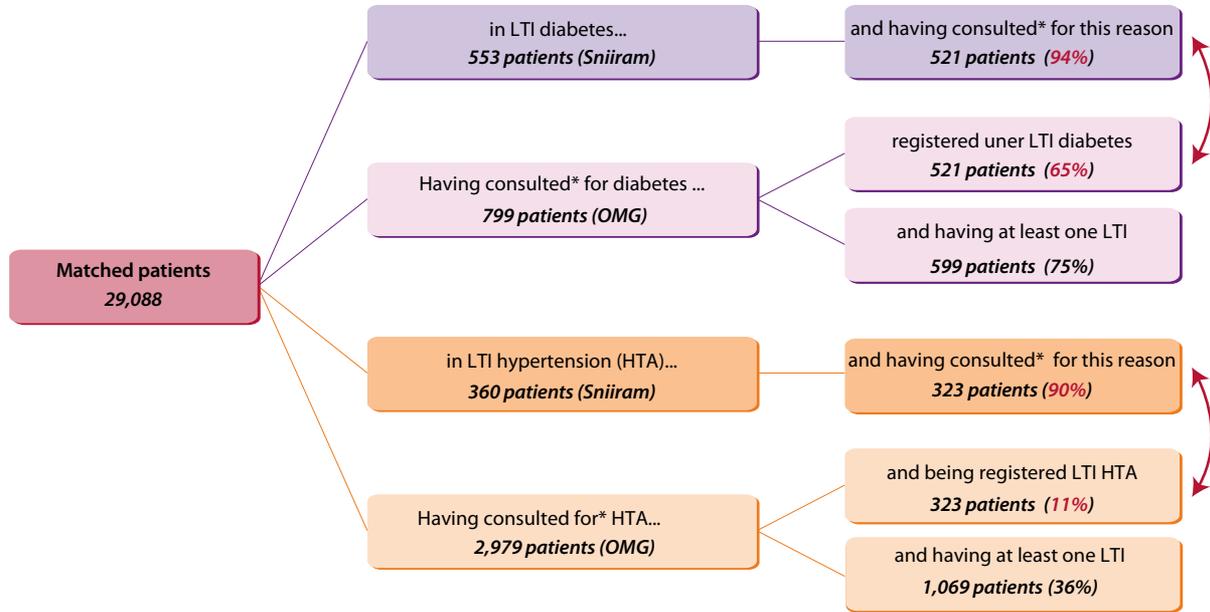
having consulted for diabetes, (at least one diabetes consultation result during the course of the year), were registered as LTI diabetes patients. Once the LTI registered patients has been identified in the SNIIRAM database, those reimbursed in 2008 for at least one anti-diabetic treatment (oral anti-diabetics or insulin) were identified. 684 patients that had been reimbursed for at least one anti-diabetic medication during the course of 2008 were thus identified, and of these, 92% had diabetes coded as a consultation result. Inversely, 20% of patients with a diabetes consultation result did not receive reimbursement for an anti-diabetic treatment during the course of the year. By crosschecking LTI diabetes identification with pharmacological treatment results, 12% of patients with a consultation result coded as diabetes were not identified in the SNIIRAM database as they were neither registered under LTI nor reimbursed for an anti-diabetic treatment during the course of the year.

Hypertension, one of the main cardiovascular risk factors, affected almost 12 million people in France in 2008. The National Nutrition and Health survey (*Etude nationale Nutrition santé*) estimated its prevalence within the population aged between 18 and 74 residing in metropolitan France at 31% (ENNS, 2007). Prevention, detection and the optimal treatment of hypertension would reduce the overall cardiovascular risk factor considerably. The identification of patients at a stage where prevention on the basis of hygiene and dietary rules is still possible is thus essential.

In this study, from the paired data set, 90% (323/360) of patients registered under LTI "hypertension" consulted one of the 30 selected GPs during the course of the year 2008. However, only 11% (323/2,979) of patients with a consultation result coded for hypertension were specifically registered as such under the LTI scheme. This rate increases to 36% (1,069/2,979) when LTI as a whole is taken into account (Graph 2). This increase is undoubtedly due to the fact that registration under the LTI code for hypertension is based on a gravity criterion. The GPs request for LTI registration cannot be accepted unless a patient's high blood pressure is treated with at least two anti-hyper-

³ Consultation results, coded according to the dictionary elaborated by the SFMG and transcodable into CIM 10, correspond to all treatment carried out by the GP during a session, whether it concerns a diagnosis, a syndrome or a symptom. Only one consultation result is coded per session but a patient may have had several results during the course of a year.

Identification of diabetic and hypertensive patients matched in OMG and SNIIRAM



* At least one session in 2008 with one of the 30 selected GPs. The reason is determined from the result of the consultation.

Source: OMG-Sniiram, 2008.

Realisation: Irdes.

tensive agents. However as hypertension was deleted from the list of recognised pathologies under the LTI scheme in 2011, this type of identification will no longer be possible; it will be even more difficult to identify hypertensive patients using the SNIIRAM database alone.

The second means of identifying hypertensive patients via the SNIIRAM database is to use reimbursed medication data in 2008. Patients thus recognised as being hypertensive were reimbursed for at least two anti-hypertensive agents from two different pharmacological classes over the course of the year. Using this method we also found that 12% of patients with a hypertensive consultation result were not directly identifiable in the SNIIRAM database via the LTI hypertension code or the presence of two anti-hypertension agents from two different pharmacological classes reimbursed during the course of the year.

* * *

The probabilistic data linkage method used within the framework of this study proved effective in linking private general practice medical records with medico-administrative data. The quality of data linkage could be improved by using

a wider SNIIRAM population base (MSA and RSI beneficiaries in addition to general regime beneficiaries from 2009) and over several consecutive years which would allow matching individuals that use general medical services less than once a year. The first exploitation of data using the test base shows the potential contribution of this type of data linkage model from the point of view of at risk population targeting and estimations of morbidity handled by general medicine.

The linkage of data from private clinical practice and SNIIRAM reimbursement data opens up new avenues for research such as the comparison of GP prescriptions (medication and procedures) with reimbursed medication for patients identifiable in the SNIIRAM database, or the analysis of patient pathways with multiple concomitant pathologies and multiple prescriptions. These are among the many themes that could be used to support the development of primary care with scientifically validated and objective measurements. However, the OMG experiment launched by the French Society of General Medicine (SFMG) to provide firm scientific backing for a practice largely unknown in scientific and professional circles was interrupted in 2011 with the closure of the OMG. This issue demonstrates the need for a broader

institutional and professional portage for a project of this magnitude. The creation of a clinical data base that can be linked with SNIIRAM data, representative of the whole of France and associating between 1000 and 2000 GPs and the 1 to 2 million patients that consult them must obtain backing from professional associations, partner institutions such as the National Health Insurance and the world of health research. In this respect, the College of General Practitioners (CMG, *Collège de la médecine générale*), in its propositions for the development of research in primary care (<http://www.lecmg.fr>) as an idea to be developed. It appears coherent, if not essential, to support through research the growth of university general medicine and more generally the first contact care sector whose mission has been defined by the 2009 HSTP Law. Beyond the feasibility test conducted by the PROSPERE team, the challenge for a health strategy aimed at strengthening the primary care sector, appears to call for further investment in research on actual practices and more globally on health care services as has been done in other research domains such as the major cohorts in epidemiology (Elfe, Constances) or the bio-banks in the biomedical sciences domain. ♦

7 FOR FURTHER INFORMATION

- Bases de données Eco-Santé, Irdes. En ligne : www.ecosante.fr
- Chevreul K., Le Fur P., Renaud T., Sermet C. (2006). « Faisabilité d'un système d'information public sur la médecine de ville ». Rapport Irdes n° 535 (biblio n° 1648), octobre.
- Cnamts (2014). « Rapport charges et produits ». www.ameli.fr/fileadmin/user_upload/documents/cnamts_rapport_charges_produits_2014.pdf
- ENNS 2006 (2007). « Situation nutritionnelle en France en 2006 selon les indicateurs d'objectif et les repères du Programme national nutrition santé (PNNS) » InVs, www.invs.sante.fr/publications/2007/nutrition_enns/
- Goldberg M., Quantin C., Guégen A., Zins M. (2012). « Bases de données médico-administratives et épidémiologie : intérêts et limites ». *Courrier des statistiques*, n° 124, mai-octobre 2008.
- Hcaam (2011). « Assurance maladie et perte d'autonomie. Contribution du HCAAM au débat sur la dépendance des personnes âgées. » www.securite-sociale.fr/IMG/pdf/hcaam_rapport_assurance_maladie_perte_autonomie.pdf
- HCSP (2009). « La prise en charge et la protection sociale des personnes atteintes de maladie chronique ». novembre.
- Institut des données de santé (IDS) (2010). « Livre blanc 2010 du Comité d'Experts : 7 propositions au service de la recherche ».
- Montaut A., Calvet L., Bouvier G., Gonzalez L. (2013). « L'appariement handicap-santé et données de l'Assurance maladie ». Drees, *Document de travail*, Série sources et méthodes, n° 39, janvier.
- Morel A., Lecoq G., Jourdain-Menninger D. (2012). « Evaluation de la prise en charge du diabète », Rapport Igas, avril www.igas.gouv.fr/IMG/pdf/RM2012-033P_Diabete_RAPPORT_TOME_I.pdf
- OMG - Observatoire de la Médecine Générale, accessible sur : omg.sfm.org/content/com/
- Païta M., Weill A. (2008). « Les personnes en affection de longue durée au 31 décembre 2007 ». Cnamts, *Points de repère* n° 20, novembre.
- Ricci P., Blotière P.-O., Weill A., Simon D., Tuppin P., Ricordeau P., Allemand H. (2010). « Diabète traité : quelles évolutions entre 2000 et 2009 en France ? » *BEH* n° 42-43, novembre.
- Silveira D.P. da, Artmann E. (2009). "Accuracy of Probabilistic Record Linkage Applied to Health Databases: Systematic Review". *Rev Saúde Pública*. oct 2009;43(5):875-882.

IRDES

INSTITUT DE RECHERCHE ET DOCUMENTATION EN ÉCONOMIE DE LA SANTÉ •
10, rue Vauvenargues 75018 Paris • Tel. : 01 53 93 43 02 • Fax : 01 53 93 43 07 •
www.Irdes.fr • Email : publications@Irdes.fr •

Director of the publication: Yann Bourgueil • **Technical senior editor:** Anne Evans • **Associate editor:** Anna Marek • **Reviewers:** Isabelle Evrard, Sylvain Pichetti •
Translator: Véronique Dandeker • **Copy Editing:** Anna Marek • **Layout compositor:** Damien Le Torrec • **ISSN :** 1283-4769.