Le repérage des maladies et du handicap dans les bases médico-administratives

Bibliographie thématique

Décembre 2019

Centre de documentation de l'Irdes

Marie-Odile Safon

Synthèses & Bibliographies

Sommaire

Problématique

Cette bibliographie a pour objectif de recenser de la littérature scientifique sur le repérage des maladies et du handicap dans les bases médico-admnistratives. Les recherches ont été effectuées sur la base Medline pour la période allant de 2010 à février 2019. Les pathologies ciblées sont les maladies relevant du système nerveux (nervous system diseases, neurodegenerative diseases), les maladies cardiovasculaires (cardiovascular diseases, stroke, ...), les cancers (neoplasms), les maladies endocriniennes (endocrine system diseases, diabetes…) et les troubles psychiatriques (mental disorders)[1]. Les références sont accompagnées de résumés et classées par ordre alphabétique d'auteurs. Cette bibliographie ne prétend pas à l'exhaustivité.

Études françaises

FOCUS : LES BASES DE DONNEES MEDICO-ADMINISTRATIVES EN FRANCE

Les sources de données en santé proviennent de cinq sources :

- Les données médico-administratives produites par l'Assurance maladie (Sniiram) et les hôpitaux (PMSI) ;
- Les données figurant dans les dossiers médicaux, à l'hôpital et en ville ;
- Les données détenues par des acteurs publics ou privés recueillies auprès de patients (essais cliniques notamment) ou de professionnels de santé ;
- Les données générées par les objets connectés, les applications mobiles, les sites web et moteurs de recherche ;
- Les données de contexte, socio-économique.

---

[1] Theaurus Mesh : https://www.ncbi.nlm.nih.gov/mesh/1000067

https://www.irdes.fr/documentation/syntheses-et-dossiers-bibliographiques.html
https://www.irdes.fr/documentation/syntheses/le-reperage-des-maladies-et-du-handicap-dans-les-bases-medico-admnistratives.pdf
https://www.irdes.fr/documentation/syntheses/le-reperage-des-maladies-et-du-handicap-dans-les-bases-medico-admnistratives.epub

---

Les bases médico-administratives en France en bref

1985 : Mise en place du PMSI (Programme de médicalisation des systèmes d'information) pour les établissements volontaires ;

1994 : Le PMSI est obligatoire pour tous les établissements de court séjour MCO (médecine chirurgie obstétrique) ;

1997 : Premières extensions du PMI au secteur privé ;

1999 : Création du Sniiram par la loi de financement de la sécurité sociale ;

2003 : Extensiondu PMSI aux SSR (soins de suite et de réadaptation) ;

2006-2007 : Extension du PMSI à la psychiatrie et à l'hospitalisation à domicile ;

2007 : Création de l'IDS (Institut des données de santé) ;

2014 : Mise en place du réseau Redsiam pour faciliter l'utilisation et la diffusion des données du Sniiram ;

2016 : Création du SNDS (Sytème national des données de santé) par la Loi n° 2016-41 du 26 janvier 2016 de modernisation de notre système de santé ;

2017 : L'IDS devient l'INDS (Institut national des données de santé) ;

2018 : Préfiguration du Health data hub sur l'analyse des données de santé et l'utilisation de l'intelligence artificielle ;

2019/02 : Lancement des appels à projet pour le Health data hub ;

2019 : l'article n° 41 de la Loi n° 2019-774 du 24 juillet 2019 relative à l'organisation et à la transformation du système de santé instaure une plate-forme des données de santé (Health data hub), qui se substitue à l'Institut national des données de santé (INDS). Cette base sécurisée pourra permettre aux chercheurs d'exploiter des données de santé élargies aux données cliniques, et plus seulement médico-administratives… avec la mention « d'intérêt public »[2].

BASES DE L'ASSURANCE MALADIE

Baujat, G., Choquet, R., Bouee, S., et al. (2017). "Prevalence of fibrodysplasia ossificans progressiva (FOP) in France: an estimate based on a record linkage of two national databases." Orphanet J Rare Dis **12**(1): 123.

> BACKGROUND: Fibrodysplasia ossificans progressiva (FOP) is a rare, severely disabling, and life-shortening genetic disorder that causes the formation of heterotopic bone within soft connective tissue. Previous studies found that the FOP prevalence was about one in every two million lives. The aim of this study is to estimate the FOP prevalence in France by probabilistic record-linkage of 2 national databases: 1) the PMSI (Programme de medicalisation des systemes d'information), an administrative database that records all hospitalization activities in France and 2) CEMARA, a registry database developed by the French Centres of Reference for Rare Diseases. RESULTS: Using a capture-recapture methodology to adjust the crude number of patients identified in both data sources, 89 FOP patients were identified, which results in a prevalence of 1.36 per million inhabitants (CI95% = [1.10; 1.68]). FOP patients' mean age was 25 years, only 14.9% were above 40 years, and 53% of them were males. The first symptoms - beside toe malformations- occurred after birth for 97.3% of them. Mean age at identified symptoms was 7 years and above 18 years for only 6.9% of patients. Mean age at diagnosis was 10 years, and above 18 years for 14.9% of the patients. FOP patients were distributed across France. CONCLUSIONS: Despite the challenge of ascertaining patients with rare diseases, we report a much higher prevalence of FOP in France than in previous studies elsewhere. We suggest that efforts to identify patients and confirm the diagnosis of FOP should be reinforced and extended at both national and European level.

Billionnet, C., Alla, F., Berigaud, E., et al. (2017). "Identifying atrial fibrillation in outpatients initiating oral anticoagulants based on medico-administrative data: results from the French national healthcare databases." Pharmacoepidemiol Drug Saf **26**(5): 535-543.

---

[2] Arrêté du 29 novembre 2019 portant approbation d'un avenant à la convention constitutive du groupement d'intérêt public « Institut national des données de santé » portant création du groupement d'intérêt public « Plateforme des données de santé »

PURPOSE: Identifying atrial fibrillation (AF) in outpatients treated with oral anticoagulants (OACs) from claims databases is challenging when the outpatient indication is not available, as OACs are also prescribed for deep vein thrombosis/pulmonary embolism (DVT/PE) that may be treated in the ambulatory setting. An algorithm was developed to identify AF in outpatients initiating OAC from medico-administrative data. METHODS: Among patients initiating OAC in 2013 in the French healthcare databases, those treated for orthopaedic indications were excluded. Patients with a history of AF or DVT/PE directly identified from available medical data, mainly hospital discharge diagnoses, were considered to be 'confirmed AF or DVT/PE patients'. Demographics of these patients and their healthcare utilization data prior to OAC initiation were then included in a logistic regression model discriminating AF versus DVT/PE indications. The final model selected, comparing c-index, provided an algorithm identifying AF from among initially unclassified patients assumed to be either AF or DVT/PE outpatients. RESULTS: Among 256 418 patients initiating OAC, 37 388 were excluded; 61 329 AF and 59 859 DVT/PE patients were directly identified, leaving 88 488 unclassified patients. The final model (c-index: 0.93) included demographics, cardiologist prescriber, hospitalization for stroke, use of antiarrhythmics/beta-blockers/antihypertensive drugs and undergoing a Holter/echocardiography procedure, thyroid function tests, but no D-dimer tests. With a specificity of 95% (sensitivity: 65%), 41% of the unclassified patients were assumed to be AF outpatients. Similar results were obtained on 250 159 new users in 2014. CONCLUSION: This algorithm combining inpatient and outpatient claims data performed relatively well to identify AF outpatients initiating OAC. Copyright (c) 2017 John Wiley & Sons, Ltd.

Blin, P., Dureau-Pournin, C., Foubert-Samier, A., et al. (2015). "Parkinson's disease incidence and prevalence assessment in France using the national healthcare insurance database." Eur J Neurol **22**(3): 464-471.

BACKGROUND AND PURPOSE: The incidence and prevalence of Parkinson's disease are important for public health planning yet there is a lack of representative, up-to-date estimations for France. METHODS: For this cross-sectional study, subjects with suspected Parkinson's were identified in the EGB database, a 1/97 random sample of the national healthcare insurance database, linked to the national hospital-discharge summary database. Incidence and prevalence were estimated using a specific definition that included those with a diagnosis (hospitalization or listed as a long-term chronic disease for full reimbursement) and a sensitive definition that also included those with an indicative drug reimbursement profile. Estimations were extrapolated to the national population, standardizing on age and gender. RESULTS: According to either the specific or the sensitive definitions, the annual incidence of Parkinson's disease during the study period was respectively 36 and 49 per 100,000 person-years and prevalence in 2010 was 308-410 per 100,000 persons in the population as a whole. According to the age groups 55-64, 65-74, 75-84 and >/=85 years incidence was respectively 33-46, 139-172, 301-363 and 442-560 per 100,000 person-years amongst men and 32-55, 81-117, 203-270 and 251-313 per 100,000 person-years amongst women. The 2010 prevalence stratified by the same age groups was 293-376, 898-1161, 2524-3011 and 3760-4578 per 100,000 persons amongst men and 199-351, 618-889, 1910-2433 and 2504-3263 per 100,000 persons amongst women. CONCLUSIONS: The specific and sensitive definitions of disease bracket the true values; the relatively small range indicates that the current study provides good estimations of incidence and prevalence of Parkinson's disease for recent years in France.

Blin, P., Dureau-Pournin, C., Lassalle, R., et al. (2017). "Outcomes in patients after myocardial infarction similar to those of the PEGASUS-TIMI 54 trial: A cohort study in the French national claims database." Br J Clin Pharmacol **83**(9): 2056-2065.

AIMS: The present study aims to describe real-life outcomes in stable patients after-myocardial infarction (MI) similar to those in the PEGASUS-TIMI 54 trial (PEGASUS), which found long-term benefits of ticagrelor in patients with a history of MI. METHODS: One-year event-free post-MI patients were identified in the French claims database representative 1/97 sample (2005-2010) and followed for up to 3 years. A PEGASUS-like (PL) population included patients with age >/= 65 years, or age >/= 50 and diabetes, renal dysfunction or prior MI, without stroke, end-stage renal failure or oral anticoagulation. Outcomes were: a composite of all-cause death or hospital admission for MI or

stroke; individual events; major bleeding. RESULTS: There were 1585 post-MI patients totalling 3926 person-years including 865 PL patients (2114 PY); 68% were male; mean age was 66 (standard deviation 15) in post-MI, 74 (10) in PL. Outcomes per 100 person-years [95% confidence interval] were, respectively, in post-MI and PL 6.3 [5.6-7.1] and 7.8 [6.7-8.9] for the composite outcome; 5.1 [4.4-5.8] and 6.5 [5.5-7.6] for death; 1.0 [0.7-1.3] and 1.0 [0.6-1.4] for MI; 0.6 [0.4-0.9] and 0.9 [0.5-1.2] for stroke; 1.3 [0.9-1.6] and 1.4 [0.9-1.9] for major bleeding. Event rates were stable over the 3 study years. Placebo patients in the PEGASUS-TIMI54 Study were younger, more often male and had lower event rates, especially for all-cause death and major bleeding. CONCLUSIONS: Patients selected using the criteria described in PEGASUS were older with more comorbidities, resulting in higher all-cause death and bleeding rates, but similar MI recurrence rates.

Bousquet, P. J., Caillet, P., Coeuret-Pellicer, M., et al. (2017). "[Using cancer case identification algorithms in medico-administrative databases: Literature review and first results from the REDSIAM Tumors group based on breast, colon, and lung cancer]." Rev Epidemiol Sante Publique **65 Suppl 4**: S236-s242.

BACKGROUND: The development and use of healthcare databases accentuates the need for dedicated tools, including validated selection algorithms of cancer diseased patients. As part of the development of the French National Health Insurance System data network REDSIAM, the tumor taskforce established an inventory of national and internal published algorithms in the field of cancer. This work aims to facilitate the choice of a best-suited algorithm. METHOD: A non-systematic literature search was conducted for various cancers. Results are presented for lung, breast, colon, and rectum. Medline, Scopus, the French Database in Public Health, Google Scholar, and the summaries of the main French journals in oncology and public health were searched for publications until August 2016. An extraction grid adapted to oncology was constructed and used for the extraction process. RESULTS: A total of 18 publications were selected for lung cancer, 18 for breast cancer, and 12 for colorectal cancer. Validation studies of algorithms are scarce. When information is available, the performance and choice of an algorithm are dependent on the context, purpose, and location of the planned study. Accounting for cancer disease specificity, the proposed extraction chart is more detailed than the generic chart developed for other REDSIAM taskforces, but remains easily usable in practice. CONCLUSIONS: This study illustrates the complexity of cancer detection through sole reliance on healthcare databases and the lack of validated algorithms specifically designed for this purpose. Studies that standardize and facilitate validation of these algorithms should be developed and promoted.

Carriere, N., Verloop, D., Dupont, C., et al. (2017). "Descriptive study of the parkinsonian population in the north of France: Epidemiological analysis and healthcare consumption." Rev Neurol (Paris) **173**(6): 396-405.

BACKGROUND: The "neurodegenerative diseases plan" under elaboration for the Hauts-de-France region requires better knowledge of the patient population and care pathways. In France, the prevalence of Parkinson's disease (PD) has been estimated from cohorts to be about 1-3 per 1000 inhabitants, but exhaustive data are scarce for the general population. The purpose of this study was to evaluate the prevalence of PD in the Hauts-de-France region and to assess PD-related healthcare consumption. METHOD: A descriptive study was conducted to identify the parkinsonian population in the Hauts-de-France region (including the administrative districts of Pas-de-Calais and Picardie) for the year 2014. Parkinsonian patients were identified from health insurance fund reimbursement data using the following criteria: (i) reimbursement for a PD-specific medication; (ii) attribution of long-duration disease status coded as PD; (iii) hospital stay with PD diagnosis in the standard discharge report contained in the French medico-economic database on hospital activity (PMSI). RESULTS: The raw prevalence of PD in the region was 5.03 per 1000 inhabitants aged 20 years and older. The standardized prevalence by health territory ranged from 4.0 to 9.0 per 1000 inhabitants aged 20 years and older. During the 1-year study period, 33.5% of patients had a neurology consultation, 57.1% attended a physiotherapy session, and 7.7% received speech therapy. Most of patients (79.6%) were treated with levodopa, sometimes in combination with a catechol-O-methyl transferase inhibitor (14.4%). Dopaminergic agonists were prescribed in 33.5% of cases. A neuroleptic was prescribed for 6.9% of the population (clozapine for 25.9%). CONCLUSION: The prevalence of PD is high in the Hauts-

de-France region with a heterogeneous distribution by health territory. Neurology consultations were attended by a minority of patients in 2014. This work provides perspectives for necessary improvement in specialized care for this disease, both in terms of follow-up consultations and home care.

Chan Chee, C., Chin, F., Ha, C., et al. (2017). "Use of medical administrative data for the surveillance of psychotic disorders in France." BMC Psychiatry **17**(1): 386.

BACKGROUND: Psychotic disorders are among the most severe psychiatric disorders that have great effects on the individuals and the society. For surveillance of chronic low prevalence conditions such as psychotic disorders, medical administrative databases can be useful due to their large coverage of the population, their continuous availability and low costs with possibility of linkage between different databases. The aims of this study are to identify the population with psychotic disorders by different algorithms based on the French medical administrative data and examine the prevalence and characteristics of this population in 2014. METHODS: The health insurance system covers the entire population living in France and all reimbursements of ambulatory care in private practice are included in a national health insurance claim database, which can be linked with the national hospital discharge databases. Three algorithms were used to select most appropriately persons with psychotic disorders through data from hospital discharge databases, reimbursements for psychotropic medication and full insurance coverage for chronic and costly conditions. RESULTS: In France in 2014, estimates of the number of individuals with psychotic disorders were 469,587 (54.6% males) including 237,808 with schizophrenia (63.6% males). Of those, 77.0% with psychotic disorders and 70.8% with schizophrenia received exclusively ambulatory care. Prevalence rates of psychotic disorders were 7.4 per 1000 inhabitants (8.3 in males and 6.4 in females) and 3.8 per 1000 inhabitants (4.9 in males and 2.6 in females) for schizophrenia. Prevalence of psychotic disorders reached a maximum of 14 per 1000 in males between 35 and 49 years old then decreased with age while in females, the highest rate of 10 per 1000 was reached at age 50 without decrease with advancing age. No such plateau was observed in schizophrenia. DISCUSSION: This study is the first in France using an exhaustive sample of medical administrative data to derive prevalence rates for psychotic disorders. Although only individuals in contact with healthcare services were included, the rates were congruent with reported estimates from systematic reviews. The feasibility of this study will allow the implementation of a national surveillance of psychotic disorders essential for healthcare management and policy planning.

Doat, S., Samson, S., Fagot-Campagna, A., et al. (2016). "Estimation of breast, prostate, and colorectal cancer incidence using a French administrative database (general sample of health insurance beneficiaries)." Rev Epidemiol Sante Publique **64**(3): 145-152.

AIM: The aim of this study was to compare incidence of breast, prostate, and colorectal cancer incidence estimated from a French administrative database with the incidences estimated from the cancer registry data. MATERIALS AND METHODS: A cohort of 426,410 people included in the general sample of health insurance beneficiaries (EGB) database as of January 1, 2007, was constituted. Several algorithms were developed to estimate cancer incidence between 2008 and 2012 using principal diagnosis (PD) of hospital discharge data (medical information systems program [PMSI]) and/or long-term disease (LTD) and together with a procedure necessary for histological diagnosis and indicating initial disease management. The incidence rates obtained were compared with those from the registry data using the standardized incidence ratio (SIR). RESULTS: The algorithm taking into account LTD and PD in the PMSI and the mandatory presence of a marker procedure provided estimates close to those from the registry data for breast cancer (SIR: 1.12 [1.07-1.18]) and colorectal cancer (SIR: 0.94 [0.88-1.02] in men and SIR: 0.93 [0.86-1.01] in women). For prostate cancer, taking into account specific procedures and drugs in addition to LTD and PD in the PMSI enhanced the estimation of incidence (SIR: 1.03 [0.98-1.08]). CONCLUSION: The PMSI together with reimbursement data (LTD, procedures, drugs) provided estimates of breast, prostate, and colorectal cancer incidence, at a national level, comparable to those from the cancer registry data.

Espagnacq, M., Penneau, A., Regaert, C. et Sermet, C. (2019). "Validation d'un algorithme de repérage des limitations visuelles dans les données du SNDS." Rev Epidemiol Sante Publique **67**: S85.
http://www.sciencedirect.com/science/article/pii/S0398762019300434

> Introduction Les données sur les personnes en situation de handicap sont essentiellement issues d'enquêtes en population générale ou ciblée par type de handicap ou de pathologie. Ces enquêtes coûteuses et peu fréquentes ne permettent pas de répondre à l'ensemble des besoins de connaissance dans ce domaine. Afin de combler ce manque, nous proposons une analyse de faisabilité du repérage des situations de handicap, approchées par les limitations fonctionnelles, à partir des données du SNDS (DCIR et PMSI). Nous présentons pour cette communication la méthodologie et les résultats du repérage des limitations visuelles. Méthodes Dans un premier temps, nous avons construit un algorithme permettant d'identifier les personnes souffrant d'une limitation visuelle à partir de différents traceurs présents dans le SNDS (motifs d'exonération, actes médicaux, rééducation, aides techniques et matériel médical, etc.). Grâce à l'appariement des données du SNDS à l'enquête Handicap Santé 2008, il est possible dans un deuxième temps de valider l'algorithme en comparant les résultats de l'algorithme aux limitations visuelles déclarées par les individus enquêtés. Cette enquête permet d'identifier de plusieurs façons les personnes qui ont des limitations visuelles (soit par la déficience soit le recours aux aides technique soit la déclaration de gêne pour voir). Résultats Qu'il s'agisse de l'enquête déclarative ou des données du SNDS, le nombre de personnes de moins de 70 ans avec d'importants problèmes visuels est très rare (moins de 5 % de la population). Discussion/Conclusion La comparaison des deux sources a permis de repérer les populations qui sont effectivement identifiables dans les données du SNDS et celles qui ne peuvent pas l'être.

Fautrel, B., Cukierman, G., Joubert, J. M., et al. (2016). "Healthcare service utilisation costs attributable to rheumatoid arthritis in France: Analysis of a representative national claims database." Joint Bone Spine **83**(1): 53-56.

> OBJECTIVE: To estimate healthcare service utilisation costs of patients with rheumatoid arthritis in France and to estimate the fraction of these costs attributable to RA. METHOD: The "Echantillon generaliste des beneficiaires" (EGB) is a 1/97 random sample of the main national claims database covering the French population. A cohort of patients with rheumatoid arthritis was constituted of all adults benefiting from full coverage for rheumatoid arthritis (ICD-10 M05-06) on 1st january 2009. A control group matched for age and gender was identified. Health expenditures were assessed from the payer's perspective for the year 2010. RESULTS: The annual per capita reimbursed total health expenditure was euro6,404 in 2010, an amount around two times higher than in the control group euro3,095 (P<0.0001). The main contributors to this extra cost were outpatient care (+euro2,407; 72.7%), including medication (+euro1,686; 50.0%), and inpatient care (+euro903; 27.3%). Patients treated by biological agents generated an age-adjusted per capita annual expenditure about three times higher than untreated patients (euro15,757 versus euro4,640). CONCLUSION: Only half of medical expenditure by patients with rheumatoid arthritis is attributable to their disease and use of biological agents has become a major driver of cost.

Filipovic-Pierucci, A., Samson, S., Fagot, J. P., et al. (2017). "Estimating the prevalence of depression associated with healthcare use in France using administrative databases." BMC Psychiatry **17**(1): 1.

> BACKGROUND: Quantitative indicators are needed in order to define priorities, plan policies and evaluate public health interventions in mental health. The aim of this study was to assess the contribution of a large and exhaustive French national administrative database to study and monitor treated depression by comparing the prevalence and characteristics of the population using significant healthcare resources for depression as identified by different estimation methods and sources and to discuss the advantages and drawbacks of these methods. METHODS: This study included the French population covered by the main health insurance scheme in 2012 (Regime general, 86% of the insured French population). Data were extracted from the French health insurance claim database (SNIIRAM), which contains information on all reimbursements, including treatments and hospital stays in France. The following distinct sources of the SNIIRAM were used to select persons with depression: diagnoses

of long-term or costly conditions, data from national hospital claims and data concerning all national health insurance reimbursements for drugs. RESULTS: In 2012, we included 58,753,200 individuals covered by the main health insurance scheme; 271,275 individuals had full coverage for depression; 179,470 individuals had been admitted to a psychiatric hospital and 66,595 individuals admitted to a general hospital with a diagnosis of depression during a 2-year timeframe and 144,670 individuals had more than three reimbursements for antidepressants during the study year (with a history of hospitalisation for depression during the past 5 years). Only 16% of individuals were selected by more than one source. CONCLUSIONS: We propose an algorithm that includes persons recently hospitalised for depression, or with a history of hospitalisation for depression and still taking antidepressants, or with full coverage for depression as a specific long-term or costly condition, yielding a prevalence estimate of 0.93% or 544,105 individuals. Changes in the case selection methodology have major consequences on the frequency count and characteristics of the selected population, and consequently on the conclusions that can be drawn from the data, emphasizing the importance of defining the characteristics of the target population before the study in order to produce relevant results.

Fosse-Edorh, S., Rigou, A., Morin, S., et al. (2017). "[Algorithms based on medico-administrative data in the field of endocrine, nutritional and metabolic diseases, especially diabetes]." <u>Rev Epidemiol Sante Publique</u> **65 Suppl 4**: S168-s173.

BACKGROUND: Medico-administrative databases represent a very interesting source of information in the field of endocrine, nutritional and metabolic diseases. The objective of this article is to describe the early works of the Redsiam working group in this field. METHODS: Algorithms developed in France in the field of diabetes, the treatment of dyslipidemia, precocious puberty, and bariatric surgery based on the National Inter-schema Information System on Health Insurance (SNIIRAM) data were identified and described. RESULTS: Three algorithms for identifying people with diabetes are available in France. These algorithms are based either on full insurance coverage for diabetes or on claims of diabetes treatments, or on the combination of these two methods associated with hospitalizations related to diabetes. Each of these algorithms has a different purpose, and the choice should depend on the goal of the study. Algorithms for identifying people treated for dyslipidemia or precocious puberty or who underwent bariatric surgery are also available. CONCLUSION: Early work from the Redsiam working group in the field of endocrine, nutritional and metabolic diseases produced an inventory of existing algorithms in France, linked with their goals, together with a presentation of their limitations and advantages, providing useful information for the scientific community. This work will continue with discussions about algorithms on the incidence of diabetes in children, thyroidectomy for thyroid nodules, hypothyroidism, hypoparathyroidism, and amyloidosis.

Foulon, S., Maura, G., Dalichampt, M., et al. (2017). "Prevalence and mortality of patients with multiple sclerosis in France in 2012: a study based on French health insurance data." <u>J Neurol</u> **264**(6): 1185-1192.

Data on the prevalence of multiple sclerosis (MS) in France are scarce. National and regional updated estimates are needed to better plan health policies. In this nationwide study, we provided estimates of the prevalence of MS in France in 2012 and mortality rate in 2013. MS cases were identified in the French national health insurance database (SNIIRAM-PMSI) using reimbursement data for disease-modifying treatment, long-term disease status for MS, disability pension for MS, and hospitalisation for MS (MS ICD-10 code: G35). We identified 99,123 MS cases, corresponding to an overall crude prevalence rate of 151.2 per 100,000 inhabitants [95% confidence interval (CI) 150.3-152.2]: 210.0 per 100,000 in women (95% CI 208.4-211.5) and 88.7 per 100,000 in men (95% CI 87.6-89.7). The overall prevalence rate was 155.6 per 100,000 inhabitants (95% CI 154.7-156.6) after standardization on the 2013-European population. We observed a prevalence gradient with a higher prevalence (190-200 per 100,000) in North-Eastern France and a lower prevalence in Southern and Western France (126-140). The crude mortality rate in 2013 was 13.7 per 1,000 MS cases (11.4 in women and 20.3 in men). The standardized mortality ratio was 2.56 (95% CI 2.41-2.72). Our results revise upwards the estimation of MS prevalence in France and confirm the excess mortality of MS patients compared to the general population.

Fuentes, S., Cosson, E., Mandereau-Bruno, L., et al. (2018). "Identifying diabetes cases in health administrative databases: a validation study based on a large French cohort." Int J Public Health.

> OBJECTIVES: In the French national health insurance information system (SNDS) three diabetes case definition algorithms are applied to identify diabetic patients. The objective of this study was to validate those using data from a large cohort. METHODS: The CONSTANCES cohort (Cohorte des consultants des Centres d'examens de sante) comprises a randomly selected sample of adults living in France. Between 2012 and 2014, data from 45,739 participants recorded in a self-administrated questionnaire and in a medical examination were linked to the SNDS. Two gold standards were defined: known diabetes and pharmacologically treated diabetes. Sensitivity, specificity, positive and negative predictive values (PPV, NPV) and kappa coefficients (k) were estimated. RESULTS: All three algorithms had specificities and NPV over 99%. Their sensitivities ranged from 73 to 77% in algorithm A, to 86 and 97% in algorithm B and to 93 and 99% in algorithm C, when identifying known and pharmacologically treated diabetes, respectively. Algorithm C had the highest k when using known diabetes as the gold standard (0.95). Algorithm B had the highest k (0.98) when testing for pharmacologically treated diabetes. CONCLUSIONS: The SNDS is an excellent source for diabetes surveillance and studies on diabetes since the case definition algorithms applied have very good test performances.

Gallini, A., Moisan, F., Maura, G., et al. (2017). "[Identification of neurodegenerative diseases in administrative databases in France: A systematic review of the literature]." Rev Epidemiol Sante Publique **65 Suppl 4**: S183-s197.

> BACKGROUND: Given the health, social and economic burden of neurodegenerative diseases (ND), the development of epidemiologic studies is required. Administrative databases, such as the French national health insurance database (SNIIRAM) could represent an opportunity for researchers. ND could be presumed from drug reimbursement data, hospital stays or registration of a chronic condition. The aim of this study was to describe, in French administrative databases, algorithms used to identify Alzheimer's disease and associated disorders (ADAD), Parkinson's disease and associated disorders (PDAD), multiple sclerosis (MS), and amyotrophic lateral sclerosis (ALS). METHODS: A systematic literature review was performed in Medline and gray literature through December 31th, 2015. French studies focusing on ADAD, PDAD, MS or ALS as a primary health outcome, conducted among one of the SNIIRAM data sources (outpatient reimbursements, chronic condition registration, hospital discharge) were included. RESULTS: Thirty-four studies were included (ADAD, n=18, PDAD, n=9, MS, n=4, ALS, n=3), leading to 36 algorithms. For each studied ND, there was an important variability in the algorithms, concerning (i) the type of criteria used (administrative database versus multi-source systems); (ii) the number of criteria used; (iii) the definition used for each criteria. The extent and level of drug exposure highly varied. Identification through hospitalizations showed variations in terms of type of stay (short stay, long-term stay, psychiatric ward...), extent of diagnosis codes used, diagnosis type (principal, related, associated diagnosis) and period used. A validation study was conducted for 2 out of 36 algorithms (PDAD), and criteria completeness was estimated for 3 algorithms (MS, ALS). CONCLUSION: Despite the increase in ND identification among French administrative databases, few algorithms have been validated. Validation studies should be encouraged.

Kab, S., Moisan, F., Preux, P. M., et al. (2017). "Nationwide incidence of motor neuron disease using the French health insurance information system database." Amyotroph Lateral Scler Frontotemporal Degener **18**(5-6): 426-433.

> OBJECTIVE: There are no estimates of the nationwide incidence of motor neuron disease (MND) in France. We used the French health insurance information system to identify incident MND cases (2012-2014), and compared incidence figures to those from three external sources. METHODS: We identified incident MND cases (2012-2014) based on three data sources (riluzole claims, hospitalisation records, long-term chronic disease benefits), and computed MND incidence by age,

gender, and geographic region. We used French mortality statistics, Limousin ALS registry data, and previous European studies based on administrative databases to perform external comparisons. RESULTS: We identified 6553 MND incident cases. After standardisation to the United States 2010 population, the age/gender-standardised incidence was 2.72/100,000 person-years (males, 3.37; females, 2.17; male:female ratio = 1.53, 95% CI1.46-1.61). There was no major spatial difference in MND distribution. Our data were in agreement with the French death database (standardised mortality ratio = 1.01, 95% CI = 0.96-1.06) and Limousin ALS registry (standardised incidence ratio = 0.92, 95% CI = 0.72-1.15). Incidence estimates were in the same range as those from previous studies. CONCLUSIONS: We report French nationwide incidence estimates of MND. Administrative databases including hospital discharge data and riluzole claims offer an interesting approach to identify large population-based samples of patients with MND for epidemiologic studies and surveillance.

Lefeuvre, D., Le Bihan-Benjamin, C., Pauporte, I., et al. (2017). "French Medico-Administrative Data to Identify the Care Pathways of Women With Breast Cancer." Clin Breast Cancer **17**(4): e191-e197.

BACKGROUND: Study of the care pathways is an important topic for care planning, as well as to observe guidelines application. This study aimed to describe care pathways and the period of time between treatments of women with breast cancer (BC), at a population level. MATERIALS AND METHODS: Women with in situ, local and regional BC who were hospitalized and newly treated in 2012 were included and followed for 1 year. Care pathways were described, focusing on surgery (partial mastectomy [PM], total mastectomy [TM]), chemotherapy, and radiotherapy. The periods of time between treatments were measured and compared with the guidelines. RESULTS: The study involved 52,128 women. The most common care pathways among the 2845 women with in situ BC were PM-radiotherapy (46.7%) and TM (28.5%). Among the 41,470 women with local BC, they were: PM-radiotherapy (44.8%) or PM-chemotherapy-radiotherapy (16.0%). The 7813 women with regional BC had similar care pathways, although chemotherapy was given more frequently (73%). The periods of time between surgery and chemotherapy were in accordance with the guidelines for 98% of the women; those between surgery and radiotherapy were affected by adjuvant chemotherapy. Finally, the time between chemotherapy and radiotherapy was longer than recommended for 40% of the women. CONCLUSION: The French medicoadministrative databases allow the study, at a national population level, of the care pathways and periods of time between treatments of women with BC according to the stage of the disease. They were close to the guidelines, although an improvement is highly necessary.

Lefeuvre, D., Rudant, J., Foulon, S., et al. (2017). "Healthcare expenditure of multiple sclerosis patients in 2013: A nationwide study based on French health administrative databases." Mult Scler J Exp Transl Clin **3**(3): 2055217317730421.

BACKGROUND: Little is known about expenditure items of multiple sclerosis (MS) patients over recent years in France. OBJECTIVE: To describe healthcare expenditure among MS patients and identify the main expenditure drivers. METHODS: All healthcare expenditure reimbursed by French National Health Insurance to MS patients in 2013 was described on the basis of nationwide health administrative databases (SNIIRAM/PMSI). Expenditure was described globally and according to age and sex. RESULTS: The average expenditure among the 90,288 MS patients included was euro11,900 per patient. Pharmacy and hospitalisation accounted for 47% and 23% of healthcare expenditure, respectively (38% and 22% of MS patients were treated with disease-modifying therapies and hospitalised overnight or longer, respectively). Average expenditure did not differ according to age. However, pharmacy expenditure decreased with age (from 71% between the ages of 20 and 29 years to 18% between the ages of 70 and 79 years), whereas hospitalisation expenditure increased with age (from 15% to 35%). Paramedical fees accounted for 2% of expenditure between the ages of 20 and 29 years and 24% between the ages of 70 and 79 years. CONCLUSION: Overall, pharmacy expenditure was the main expenditure item, which decreased with increasing age, while hospitalisation and paramedical expenditure increased with increasing age.

Maroun, R., Fleury, L., Nachbaur, G., et al. (2017). "Real-world costs and outcomes in metastatic renal cell carcinoma patients treated with targeted therapies: a cohort study from the French health insurance database." Curr Med Res Opin **33**(10): 1755-1762.

OBJECTIVES: The objective of this study was to describe treatment patterns, survival, healthcare use and costs in patients with metastatic renal cell carcinoma (mRCC) in a real-world setting. RESEARCH DESIGN AND METHODS: We used the National Health Insurance (NHI) claims database for the Ile-de-France region to perform a retrospective cohort analysis of patients with mRCC treated by a first-line targeted therapy. Treatment naive patients were identified combining the 10th revision of the International Classification of Diseases (ICD-10) codes (C64 & C77-C79) and a first prescription of targeted therapies. Descriptive analyses were performed on treatment patterns and patients' characteristics. Progression free survival (PFS) and overall survival (OS) were determined using Kaplan-Meier actuarial survival analysis. All healthcare resource use and costs were estimated on a per patient per month (PPPM) basis (euro2016). RESULTS: A total of 327 treatment naive patients with mRCC were included. Median follow-up was 13.4 months. Sunitinib accounted for 73% of first-line treatments. The most frequently observed treatment sequence for the first two lines was sunitinib-everolimus (16%; n = 137) and for the first three lines sunitinib-everolimus-axitinib (20%; n = 49). First-line PFS for sunitinib, everolimus, pazopanib, sorafenib and other was 8.7, 6.2, 10.7, 5.7 and 11.2 months, respectively. Median OS for patients treated by first-line sunitinib, everolimus, pazopanib, sorafenib and other was respectively 14.7, 8.1, 21.1, 8.9 and 14.0 months. From the NHI's perspective, the mean PPPM was euro5546. The average PPPM in pre-progression was euro5597 compared to euro5541 beyond progression of the disease. Oral targeted therapies accounted for 53% of the total PPPM. CONCLUSION: This descriptive study showed that the economic burden of mRCC is substantial with oral targeted therapies accounting for 53% of the PPPM. OS and PFS in real life are poorer than observed in clinical trials.

Mechtouff, L., Haesebaert, J., Viprey, M., et al. (2018). "Secondary Prevention Three and Six Years after Stroke Using the French National Insurance Healthcare System Database." Eur Neurol **79**(5-6): 272-280.

BACKGROUND: Secondary prevention is inadequate in the first 2 years after stroke but what happens after that is less documented. The aim of this study was to assess the use and the adherence to preventive drugs 3 and 6 years after experiencing a transient ischemic attack (TIA) or an ischemic stroke (IS). METHODS: The population study was from the AVC69 cohort (IS or TIA admitted in an emergency or stroke unit in the Rhone area, France, for an IS or a TIA during a 7-month period). Medication use was defined as >/=1 purchase during the studied year and adherence as Continuous Measure of Medication Acquisition >/=0.8 using the French medical insurance health care funding database. RESULTS: The study population consisted of 210 patients at 3 years and 163 patients at 6 years. Medication use at 3 and 6 years was, respectively, 80.9 and 79.8% for antithrombotics, 69.1 and 66.3% for antihypertensives, 60.5 and 55.2% for statins and 48.6 and 46.6% for optimal treatment defined as the treatment achieved by the use of the 3 drugs. Adherence to each class was good at 3 years and tends to decrease at 6 years. CONCLUSIONS: More than one patient out of 2 do not use the optimal preventive treatment.

Mezaache, S., Derumeaux, H., Ferraro, P., et al. (2017). "Validation of an algorithm identifying incident primary immune thrombocytopenia in the French national health insurance database." Eur J Haematol **99**(4): 344-349.

OBJECTIVES: To evaluate the accuracy of an algorithm identifying newly diagnosed immune thrombocytopenia (ITP) patients in the French national health insurance database (SNIIRAM). METHODS: The source of data was the SNIIRAM of Midi-Pyrenees region (southwest of France, three million inhabitants). Data of patients with at least one ITP code (D69.3 code of the International Classification of Disease, version 10) were extracted between January 1, 2012, and December 31, 2014. We used an algorithm that identifies newly diagnosed primary ITPs. Medical charts of incident ITPs were reviewed. Positive predictive values (PPVs) of identification of true, incident, and primary ITP cases were estimated. RESULTS: Of the 168 patients selected, 161 were true ITP cases yielding a PPV of 95.8% (95% confidence interval-95% CI: 92.8-98.8). Among them, 128 were truly incident

according to symptom onset date and 134 according to the diagnosis date yielding PPVs of 79.5% (95% CI: 73.2-85.7) and 83.2% (95% CI: 77.4-89.0), respectively. Median time between estimated diagnosis date by the algorithm and true diagnosis date was 0 days (interquartile range: 0 to 15). CONCLUSIONS: This study showed a very good PPV of this algorithm identifying incident primary ITP patients in the SNIIRAM.

Palmaro, A., Gauthier, M., Conte, C., et al. (2017). "Identifying multiple myeloma patients using data from the French health insurance databases: Validation using a cancer registry." Medicine (Baltimore) **96**(12): e6189.

This study aimed to assess the performance of several algorithms based on hospital diagnoses and the long-term diseases scheme to identify multiple myeloma patients.Potential multiple myeloma patients in 2010 to 2013 were identified using the presence of hospital records with at least 1 main diagnosis code for multiple myeloma (ICD-10 "C90"). Alternative algorithms also considered related and associated diagnoses, combination with long-term conditions, or at least 2 diagnoses. Incident patients were those with no previous "C90" codes in the past 24 or 12 months. The sensitivity, specificity, and positive and negative predictive values (PPVs and NPVs) were computed, using a French cancer registry for the corresponding area and period as the criterion standard.Long-term conditions data extracted concerned 11,559 patients (21,846 for hospital data). The registry contained 125 cases of multiple myeloma. Sensitivity was 70% when using only main hospital diagnoses (specificity 100%, PPV 79%), 76% when also considering related diagnoses (specificity 100%, PPV 74%), and 90% with associated diagnoses included (100% specificity, 64% PPV).In relation with their good performance, selected algorithms can be used to study the benefit and risk of drugs in treated multiple myeloma patients.

Quantin, C., Collin, C., Frerot, M., et al. (2017). "[Study of algorithms to identify schizophrenia in the SNIIRAM database conducted by the REDSIAM network]." Rev Epidemiol Sante Publique **65 Suppl 4**: S226-s235.

BACKGROUND: The aim of the REDSIAM network is to foster communication between users of French medico-administrative databases and to validate and promote analysis methods suitable for the data. Within this network, the working group "Mental and behavioral disorders" took an interest in algorithms to identify adult schizophrenia in the SNIIRAM database and inventoried identification criteria for patients with schizophrenia in these databases. METHODS: The methodology was based on interviews with nine experts in schizophrenia concerning the procedures they use to identify patients with schizophrenia disorders in databases. The interviews were based on a questionnaire and conducted by telephone. RESULTS: The synthesis of the interviews showed that the SNIIRAM contains various tables which allow coders to identify patients suffering from schizophrenia: chronic disease status, drugs and hospitalizations. Taken separately, these criteria were not sufficient to recognize patients with schizophrenia, an algorithm should be based on all of them. Apparently, only one-third of people living with schizophrenia benefit from the longstanding disease status. Not all patients are hospitalized, and coding for diagnoses at the hospitalization, notably for short stays in medicine, surgery or obstetrics departments, is not exhaustive. As for treatment with antipsychotics, it is not specific enough as such treatments are also prescribed to patients with bipolar disorders, or even other disorders. It seems appropriate to combine these complementary criteria, while keeping in mind out-patient care (every year 80,000 patients are seen exclusively in an outpatient setting), even if these data are difficult to link with other information. Finally, the experts made three propositions for selection algorithms of patients with schizophrenia. CONCLUSION: Patients with schizophrenia can be relatively accurately identified using SNIIRAM data. Different combinations of the selected criteria must be used depending on the objectives and they must be related to an appropriate length of time.

Roxburgh, R. H., Seaman, S. R., Masterman, T., et al. (2005). "Multiple Sclerosis Severity Score: using disability and disease duration to rate disease severity." Neurology **64**(7): 1144-1151.

BACKGROUND: There is no consensus method for determining progression of disability in patients with multiple sclerosis (MS) when each patient has had only a single assessment in the course of the disease. METHODS: Using data from two large longitudinal databases, the authors tested whether

cross-sectional disability assessments are representative of disease severity as a whole. An algorithm, the Multiple Sclerosis Severity Score (MSSS), which relates scores on the Expanded Disability Status Scale (EDSS) to the distribution of disability in patients with comparable disease durations, was devised and then applied to a collection of 9,892 patients from 11 countries to create the Global MSSS. In order to compare different methods of detecting such effects the authors simulated the effects of a genetic factor on disability. RESULTS: Cross-sectional EDSS measurements made after the first year were representative of overall disease severity. The MSSS was more powerful than the other methods the authors tested for detecting different rates of disease progression. CONCLUSION: The Multiple Sclerosis Severity Score (MSSS) is a powerful method for comparing disease progression using single assessment data. The Global MSSS can be used as a reference table for future disability comparisons. While useful for comparing groups of patients, disease fluctuation precludes its use as a predictor of future disability in an individual.

Thurin, N., Rouyer, M., Gross-Goupil, M., Soulié, M., Roumiguié, M., Le Moulec, S., De Beaucoudrey, L., Droz-Perroteau, C. et Blin, P. (2019). "Validation d'un algorithme complexe dans le SNDS. Exemple du cancer de la prostate résistant à la castration et métastatique." Rev Epidemiol Sante Publique **67**: S75. http://www.sciencedirect.com/science/article/pii/S039876201930015X

Introduction Un algorithme d'identification des cas de cancer de la prostate résistant à la castration et métastatique (mCRPC) a été construit dans le Système national des données de santé (SNDS). Cet algorithme a fait l'objet d'une étude de validation afin d'en déterminer les caractéristiques extrinsèques. Méthodes Un tirage au sort a été effectué pour sélectionner 100 cas parmi les 14 050 patients mCRPC identifiés par l'algorithme au sein du SNDS et 100 cas parmi 372 273 patients non-mCRPC. Après anonymisation, l'historique des soins médicaux sur six années a été reconstitué pour chacun de ces 200 cas à partir des données disponibles dans le SNDS (ALD, délivrances de médicaments, actes CCAM, hospitalisations, tests biologiques). Un comité formé de deux urologues et de deux oncologues répartis par binôme (urologue-oncologue) a évalué le statut mCRPC des cas tirés au sort à partir de ces informations et en aveugle des résultats de l'algorithme. Chaque binôme s'est vu attribué la relecture de 100 cas. Lors de discordance au sein d'un binôme sur un statut, le cas était revu de façon collégiale par les quatre experts. Les valeurs prédictives positives (VPP) et négatives (VPN) de l'algorithme ont été calculées par rapport aux conclusions du comité. Résultats Après confrontation des résultats de l'algorithme aux conclusions du comité, 92 des 100 cas mCRPC et 93 des 100 cas non-mCRPC ont été confirmés. Il en résulte une VPP de 0,92 et une VPN de 0,93. Discussion/Conclusion La richesse des données à disposition dans le SNDS rend possible la mise en œuvre d'algorithmes de détection de pathologies complexes, mais aussi d'études de validation via la reconstitution de dossiers médicaux anonymisés. Les résultats montrent que l'algorithme d'identification présenté permet une bonne identification des cas de mCRPC dans le SNDS. En outre, l'étude de validation a permis de détecter quelques paramètres à ajuster afin d'en optimiser les performances.

Tilleul, P., Jacot, W., Emery, C., et al. (2017). "Management and cost analysis of cancer patients treated with G-CSF: a cohort study based on the French national healthcare insurance database." J Med Econ **20**(12): 1261-1267.

OBJECTIVES: To describe the management and costs associated with G-CSF therapy in cancer patients in France. METHODS: This study analyzed a representative random population sample from the French national healthcare insurance database, focusing on 1,612 patients with hematological or solid malignancies who were reimbursed in 2013 or 2014 for at least one G-CSF treatment dispensed in a retail pharmacy. Patient characteristics and treatment costs were analyzed according to the type of cancer. Then the costs and characteristics of patients associated with the use of different G-CSF products were analyzed in the sub-set of breast cancer patients. RESULTS: The most frequent malignancies in the database population were breast cancer (23.3%), hematological malignancies

(22.2%), and lung cancer (12.4%). The reimbursed G-CSF was pegfilgrastim in 34.1% of cases, lenograstim in 26.7%, and filgrastim in 17.9%. More than one G-CSF product was reimbursed to 21.3% of patients. The total annual reimbursed health expenses per patient, according to the type of G-CSF, were euro27,001, euro24,511, and euro20,802 for patients treated with filgrastim, lenograstim, and pegfilgrastim, respectively. Ambulatory care accounted for, respectively, 35%, 38%, and 41% of those costs. In patients with breast cancer, ambulatory care cost was euro7,915 with filgrastim, euro7,750 with lenograstim, and euro6,989 with pegfilgrastim, and the respective cost of G-CSF was euro1,733, euro1,559, and euro3,668. CONCLUSION: All available G-CSF products have been shown to be effective in cancer patients, and both daily G-CSFs and pegylated G-CSF are recommended in international guidelines. Nevertheless, this analysis of G-CSF reimbursement indicates that the choice of product can markedly affect the total cost of ambulatory care.

Tuppin, P., Bessou, A., Legeai, C., et al. (2018). "Two-year management after renal transplantation in 2013 in France: Input from the French national health system database." Nephrol Ther **14**(4): 207-216.

The objective of this study was to describe the management of patients undergoing renal transplantation in 2013 and over the following two years on the basis of healthcare consumption data. The National Health Insurance Information System was used to identify 1876 general scheme beneficiaries undergoing a first isolated renal transplantation (median age: 53 years; men 63%). Overall, 1.2% of patients died during the transplantation hospital stay (>65 years 3.3%) and 87% of patients had a functional graft at 2 years. Thirty-three percent of patients were readmitted to hospital for 1 day or longer during the first month, 73% the first year and 55% the second year. At least 10% of patients were hospitalised for antirejection treatment during the first quarter after renal transplantation, 16% the first year and 9% the second year. The first year, 32% of patients were hospitalised for renal disease (12% the second year), 14% were hospitalised for cardiovascular disease (9% the second year), 13% for infectious disease (5% the second year) and 2% for a malignant tumour (2% the second year). Almost 80% of patients consulted their general practitioner each year (almost 50% consulted every quarter). During the second year, 83% of patients were taking antihypertensives, 45% lipid-lowering drugs, 26% antidiabetic drugs, 77% tacrolimus, 18% ciclosporin, 88% mycophenolic acid and 69% corticosteroids. This study highlights the important contribution of healthcare consumption data to a better understanding of the modalities of management of renal transplant recipients in France, allowing improvement of this management in line with guidelines.

Tuppin, P., Pestel, L., Samson, S., et al. (2017). "[The human and economic burden of cancer in France in 2014, based on the Sniiram national database]." Bull Cancer **104**(6): 524-537.

INTRODUCTION: The national health insurance information system (Sniiram) can be used to estimate the national medical and economic burden of cancer. This study reports the annual rates, characteristics and expenditure of people reimbursed for cancer. METHODS: Among 57 million general health scheme beneficiaries (86% of the French population), people managed for cancer were identified using algorithms based on hospital diagnoses and full refund for long-term cancer. The reimbursed costs (euros) related to the cancer, paid off by the health insurance, were estimated. RESULTS: In 2014, 2.491 million people (4.4%) covered by the general health scheme had a cancer managed (men 1.1 million, 5.1%; women 1.3 million, 4.9%). The annual (2012-2014) average growth rate of patients was 0.8%. The spending related to the cancer was 13.5 billion: 5 billion for primary health care (drugs 2.3 billion), 7.5 billion for the hospital (drugs 1.3 billions) and 900 million for sick leave and invalidity pensions. Spending annual average growth rate (2012-2014) was 4% (drugs 2%). The rates of patients and the relative spending were 1.8% and 2.5 billion for the breast cancer (women), 1.5% and 1.0 billion for prostate cancer, 0.9% and 1.5 billion for the colon cancer, and 0.19% and 1.3 billion for lung cancer. DISCUSSION: Cancers establish one of the first groups of chronic diseases pathologies in terms of patients and spending. If the numbers of patients remain stables, the spending increases, mainly for medicines.

Tuppin, P., Riviere, S., Rigault, A., et al. (2016). "Prevalence and economic burden of cardiovascular diseases in France in 2013 according to the national health insurance scheme database." Arch Cardiovasc Dis **109**(6-7): 399-411.

> BACKGROUND: Cardiovascular diseases (CVDs) constitute the second leading cause of death in France. The Systeme national d'information interregimes de l'assurance maladie (SNIIRAM; national health insurance information system) can be used to estimate the national medical and economic burden of CVDs. OBJECTIVES: To describe the rates, characteristics and expenditure of people reimbursed for CVDs in 2013. METHODS: Among 57 million general health scheme beneficiaries (86% of the French population), people managed for CVDs were identified using algorithms based on hospital diagnoses either during the current year (acute phase) or over the previous 5 years (chronic phase) and long-term diseases. The reimbursed costs attributable to CVDs were estimated. RESULTS: A total of 3.5 million people (mean age, 71 years; 42% women) were reimbursed by the general health scheme for CVDs (standardized rate, 6.5%; coronary heart disease, 2.7%; arrhythmias/conduction disorders, 2.1%; stroke, 1.1%; heart failure, 1.1%). These frequencies increased with age and social deprivation, and were higher in Northern and Eastern France and Reunion Island. The total sum reimbursed by all schemes for CVDs was euro 15.1 billion (50% for hospital care and 43% for outpatient care [including 15% for drugs and 12% for nurses/physiotherapists]); coronary heart disease accounted for euro 4 billion, stroke for euro 3.5 billion and heart failure for euro 2.5 billion (i.e. 10% of the total expenditure reimbursed by all national health insurance schemes for all conditions). CONCLUSION: CVDs constitute the leading group in terms of numbers of patients reimbursed and total reimbursed expenditure, despite a probable underestimation of both numbers and expenditure.

Tuppin, P., Rudant, J., Constantinou, P., et al. (2017). "Value of a national administrative database to guide public decisions: From the systeme national d'information interregimes de l'Assurance Maladie (SNIIRAM) to the systeme national des donnees de sante (SNDS) in France." Rev Epidemiol Sante Publique **65 Suppl 4**: S149-s167.

> In 1999, French legislators asked health insurance funds to develop a systeme national d'information interregimes de l'Assurance Maladie (SNIIRAM) [national health insurance information system] in order to more precisely determine and evaluate health care utilization and health care expenditure of beneficiaries. These data, based on almost 66 million inhabitants in 2015, have already been the subject of numerous international publications on various topics: prevalence and incidence of diseases, patient care pathways, health status and health care utilization of specific populations, real-life use of drugs, assessment of adverse effects of drugs or other health care procedures, monitoring of national health insurance expenditure, etc. SNIIRAM comprises individual information on the sociodemographic and medical characteristics of beneficiaries and all hospital care and office medicine reimbursements, coded according to various systems. Access to data is controlled by permissions dependent on the type of data requested or used, their temporality and the researcher's status. In general, data can be analyzed by accredited agencies over a period covering the last three years plus the current year, and specific requests can be submitted to extract data over longer periods. A 1/97th random sample of SNIIRAM, the echantillon generaliste des beneficiaires (EGB), representative of the national population of health insurance beneficiaries, was composed in 2005 to allow 20-year follow-up with facilitated access for medical research. The EGB is an open cohort, which includes new beneficiaries and newborn infants. SNIIRAM has continued to grow and extend to become, in 2016, the cornerstone of the future systeme national des donnees de sante (SNDS) [national health data system], which will gradually integrate new information (causes of death, social and medical data and complementary health insurance). In parallel, the modalities of data access and protection systems have also evolved. This article describes the SNIIRAM data warehouse and its transformation into SNDS, the data collected, the tools developed in order to facilitate data analysis, the limitations encountered, and changing access permissions.

AUTRES BASES ADMINISTRATIVES

Coureau, G., Baldi, I., Saves, M., et al. (2012). "[Performance evaluation of hospital claims database for the identification of incident central nervous system tumors compared with a cancer registry in Gironde, France, 2004]." Rev Epidemiol Sante Publique **60**(4): 295-304.

> BACKGROUND: Cancer registries cover 18% of the French population. A national surveillance might be warranted for some potentially environment-related cancers such as tumors of the central nervous system (CNS) to detect abnormal incidence variations. The PMSI database provides an interesting source of comprehensive, standardized and mandatory data collected from all health facilities. The aim of this work was to develop methods to identify incident CNS tumors using the PMSI database. METHODS: A selection of patients living in Gironde was made in the 2004 PMSI database of the hospital of Bordeaux, using the CNS tumors codification. Cases were validated via the CNS primary tumor registry of Gironde taken as the reference, or medical records. Various combinations of criteria were defined and tested. RESULTS: The first selection based on diagnoses identified patients with a sensitivity of 84% and a positive predictive value (PPV) of 34%. Patients wrongly identified by the PMSI were non-incident cases (49%) or patients without a CNS tumor (45%). Patients with a tumor not identified by the PMSI had been hospitalized in 2005 (44%) or had no code for CNS tumor (42%). According to the algorithms, the sensitivity ranged from 64% to 84%, and the PPV from 34% to 69%. The best combination had a sensitivity of 67% and a PPV of 69% and was obtained with codes for CNS tumor in 2004 associated with a diagnostic or therapeutic code for persons under 70 years without code for CNS tumor in previous years or code for metastasis in 2004. CONCLUSION: According to these results, the PMSI database cannot be used alone to calculate the incidence of these complex tumors. However the PMSI database plays an important role in cancer surveillance, in combination with other information sources and the expertise of cancer registries. This role could increase with further reflection and improvement of data quality.

Lecoffre, C., de Peretti, C., Gabet, A., et al. (2017). "National Trends in Patients Hospitalized for Stroke and Stroke Mortality in France, 2008 to 2014." Stroke **48**(11): 2939-2945.

> BACKGROUND AND PURPOSE: Stroke is the leading cause of death in women and the third leading cause in men in France. In young adults (ie, <65 years old), an increase in the incidence of ischemic stroke was observed at a local scale between 1985 and 2011. After the implementation of the 2010 to 2014 National Stroke Action Plan, this study investigates national trends in patients hospitalized by stroke subtypes, in-hospital mortality, and stroke mortality between 2008 and 2014. METHODS: Hospitalization data were extracted from the French national hospital discharge databases and mortality data from the French national medical causes of death database. Time trends were tested using a Poisson regression model. RESULTS: From 2008 to 2014, the age-standardized rates of patients hospitalized for ischemic stroke increased by 14.3% in patients <65 years old and decreased by 1.5% in those aged >/=65 years. The rate of patients hospitalized for hemorrhagic stroke was stable (+2.0%), irrespective of age and sex. The proportion of patients hospitalized in stroke units substantially increased. In-hospital mortality decreased by 17.1% in patients with ischemic stroke. From 2008 to 2013, stroke mortality decreased, except for women between 45 and 64 years old and for people aged >/=85 years. CONCLUSIONS: An increase in cardiovascular risk factors and improved stroke management may explain the increase in the rates of patients hospitalized for ischemic stroke. The decrease observed for in-hospital stroke mortality may be because of recent improvements in acute-phase management.

Maire, F., Cibot, J. O., Compagne, C., et al. (2017). "Epidemiology of pancreatic cancer in France: descriptive study from the French national hospital database." Eur J Gastroenterol Hepatol **29**(8): 904-908.

> INTRODUCTION: Although indirect evidence suggests that the incidence of pancreatic adenocarcinoma has increased in the last decade, few data are available in European countries. The aim of the present study was to update the epidemiology of pancreatic cancer in France in 2014 from the French national hospital database (Programme de Medicalisation des Systemes d'Information). PATIENTS AND METHODS: All patients hospitalized for pancreatic cancer in France in 2014 in public or private institutions were included. Patient and stays (length, type of support, institutions) characteristics were

studied. The results were compared with those observed in 2010. RESULTS: A total of 13 346 (52% men, median age 71 years) new patients were treated for pancreatic cancer in 2014, accounting for a 12.5% increase compared with 2010. Overall, 22% of patients were operated on. Liver metastases were present in 60% of cases. The disease accounted for 146 680 hospital stays (+24.8% compared with 2010), 76% of which were related to chemotherapy (+32%). The average annual number and length of stay were 7 and 2.6 days, respectively. In 2014, 11 052 deaths were reported (+15.8%). CONCLUSION: Approximately 13 350 new cases of pancreatic cancer were observed in France in 2014. The increase in incidence was associated with a marked increase in hospital stays for chemotherapy.

Mercier, G., Duflos, C., Riondel, A., et al. (2018). "Admissions to intensive cardiac care units in France in 2014: A cross-sectional, nationwide population-based study." Medicine (Baltimore) **97**(40): e12677.

Geographic variation in admission to the intensive cardiac care unit (ICCU) might question about the efficiency and the equity of the healthcare system. The aim was to explain geographic variation in the rate of admission to ICCU for coronary artery disease (CAD) or heart failure (HF) in France.We conducted a retrospective study based on the French national hospital discharge database. All inpatient stays for CAD or HF with an admission to an ICCU in 2014 were included. We estimated population-based age and sex-standardized ICCU admission rates at the department level. We separately modeled the department-level admission rates for HF and CAD using generalized linear models.In all, 61,010 stays for CAD and 27,828 stays for HF had at least 1 ICCU admission. The ICCU admission rates were explained by the admission rate for CAD, by the diabetes prevalence, by the proportion of the population >75 years, and by the drive time to the ICCU.This work sheds light on the finding of substantial geographic variation in the ICCU admission rates for CAD and HF in France. This variation is explained by both the age and the health status of the population and also by the drive time to the closest ICCU for HF. Moreover, ICCU admission for HF might be more prone to unwarranted variations due to medical practice patterns.

Roubille, F., Mercier, G., Delmas, C., et al. (2017). "Description of acute cardiac care in 2014: A French nation-wide database on 277,845 admissions in 270 ICCUs." Int J Cardiol **240**: 433-437.

BACKGROUND: The Intensive Cardiac Care Unit (ICCU) has greatly evolved for decades: it no longer includes only patients with coronary artery disease (CAD). The clinical characteristics and pathological profiles of patients have markedly changed. Detailed data on the topic are critically lacking. METHODS: We present here a French nation-wide administrative database with an exhaustive description of patients admitted to ICCU throughout a whole year (2014). RESULTS: A total of 277,845 patients in 270 centers were admitted to ICCUs at least once in 2014 (exhaustive data). Median age was 71years (IQR: 59-81) and the patients were primarily male (63%). Mean ICCU stay was 2.0days (1.0-4.0). CAD patients (49.0%) represented the major group admitted, followed by patients with arrhythmias (15.2%) and heart failure (HF) (10.0%). Patients admitted with acute CAD were significantly younger (mean age 67.4 y), had better outcomes (mortality 4.0%), and shorter hospital stays (mean stay 6.7 d). Patients with HF were significantly older (mean age 75.2 y), with longer hospital stays (mean stay 12.0 d), and poorer outcomes (mortality 10.5%). CONCLUSION: We present here the largest contemporary administrative database on patients admitted to ICCUs in a developed country. CAD (mainly acute coronary syndromes) remains the primary cause of admission but the population is, by far, more complex than generally considered.

Roussot, A., Cottenet, J., Gadreau, M., et al. (2016). "The use of national administrative data to describe the spatial distribution of in-hospital mortality following stroke in France, 2008-2011." Int J Health Geogr **15**: 2.

BACKGROUND: In the context of implementing the National Stroke Plan in France, a spatial approach was used to measure inequalities in this disease. Using the national PMSI-MCO databases, we analyzed the in-hospital prevalence of stroke and established a map of in-hospital mortality rates with regard to the socio-demographic structure of the country. METHODS: The principal characteristics of patients identified according to ICD10 codes relative to stroke (in accordance with earlier validation work) were studied. A map of standardized mortality rates at the level of PMSI geographic codes was

established. An exploratory analysis (principal component analysis followed by ascending hierarchical classification) using INSEE socio-economic data and mortality rates was also carried out to identify different area profiles. RESULTS: Between 2008 and 2011, the number of stroke patients increased by 3.85%, notably for ischemic stroke in the 36-55 years age group (60% of men). Over the same period, in-hospital mortality fell, and the map of standardized rates illustrated the diagonal of high mortality extending from the north-east to the south-west of the country. The most severely affected areas were also those with the least favorable socio-professional indicators. CONCLUSIONS: The PMSI-MCO database is a major source of data on the health status of the population. It can be used for the area-by-area observation of the performance of certain healthcare indicators, such as in-hospital mortality, or to follow the implementation of the National Stroke Plan. Our study showed the interplay between social and demographic factors and stroke-related in-hospital mortality. The map derived from the results of the exploratory analysis illustrated a variety of areas where social difficulties, aging and high mortality seemed to meet. The study raises questions about access to neuro-vascular care in isolated areas and in those in demographic decline. Telemedicine appears to be the solution favored by decision makers. The aging of the population managed for stroke must not mask the growing incidence in younger people, which raises questions about the development of classical (smoking, hypertension) or new (drug abuse) risk factors.

METHODES, MODELES

Ajrouche, A., Estellat, C., De Rycke, Y., et al. (2017). "Evaluation of algorithms to identify incident cancer cases by using French health administrative databases." Pharmacoepidemiol Drug Saf **26**(8): 935-944.

Bezin, J., Duong, M., Lassalle, R., et al. (2017). "The national healthcare system claims databases in France, SNIIRAM and EGB: Powerful tools for pharmacoepidemiology." Pharmacoepidemiol Drug Saf **26**(8): 954-962.

Bousquet, P. J., Caillet, P., Coeuret-Pellicer, M., et al. (2017). "[Using cancer case identification algorithms in medico-administrative databases: Literature review and first results from the REDSIAM Tumors group based on breast, colon, and lung cancer]." Rev Epidemiol Sante Publique **65 Suppl 4**: S236-s242.

Drezen, E., Guyet, T. et Happe, A. (2018). "From medico-administrative databases analysis to care trajectories analytics: an example with the French SNDS." Fundam Clin Pharmacol **32**(1): 78-80.

Ferdynus, C. et Huiart, L. (2016). "[Technical improvement of cohort constitution in administrative health databases: Providing a tool for integration and standardization of data applicable in the French National Health Insurance Database (SNIIRAM)]." Rev Epidemiol Sante Publique **64**(4): 263-269.

Ferdynus, C. et Hulard, L. (2016). "Optimisation de la constitution de cohortes issues de bases de données médico-administratives : mise à disposition d'un algorithme pour l'intégration et la normalisation des données adapté au Système national d'information inter-régimes de l'assurance maladie (SNIIRAM)". Revue d'Epidémiologie et de Santé Publique **64**(4): 263-269.
http://www.em-consulte.com/article/1080254/alerteM

Fonteneau, L., Le Meur, N., Cohen-Akenine, A., et al. (2017). "[The use of administrative health databases in infectious disease epidemiology and public health]." Rev Epidemiol Sante Publique **65 Suppl 4**: S174-s182.

Fosse-Edorh, S., Rigou, A., Morin, S., et al. (2017). "[Algorithms based on medico-administrative data in the field of endocrine, nutritional and metabolic diseases, especially diabetes]." Rev Epidemiol Sante Publique **65 Suppl 4**: S168-s173.

Gissot, C. (2018). Les données médico-administratives de l'Assurance maladie : les utilisations par la Cnam. Santé et intelligence artificielle., Paris : CNRS éditions**:** 75-84.

Goldberg, M. et al. (2016). "L'ouverture du Système national d'information inter-régimes de l'assurance maladie (SNIIRAM) : des opportunités et des difficultés. L'expérience des cohortes Gazel et Constances." Revue d'Epidémiologie et de Santé Publique **64 ;**

Goldberg, M., Carton, M., Doussin, A., et al. (2017). "Le réseau REDSIAM (Réseau données Sniiram) – Spécial REDSIAM." Revue d'Epidémiologie et de Santé Publique **65**: S144-S148. http://www.sciencedirect.com/science/article/pii/S0398762017304480

Goldberg, M., Carton, M., Gourmelen, J., et al. (2016). "[The opening of the French national health database: Opportunities and difficulties. The experience of the Gazel and Constances cohorts]." Rev Epidemiol Sante Publique **64**(4): 313-320.

Happe, A. et Drezen, E. (2018). "A visual approach of care pathways from the French nationwide SNDS database - from population to individual records: the ePEPS toolbox." Fundam Clin Pharmacol **32**(1): 81-84.

Looten, V. (2019). "Les études portant sur les bases de données médico-administratives sont-elles reproductibles ? L'hypothèse d'une inconduite éthique en santé publique." (35): 689–692. https://www.medecinesciences.org/fr/articles/medsci/full_html/2019/08/msc190113/msc190113.html?utm_source=email_alert_medsci&utm_medium=email&utm_campaign=2019-09-18

Lukacs, B. (2018). Comment la base Sniiram-PMSI permet l'étude des pratiques en chirurgie. Santé et intelligence artificielle., Paris : CNRS éditions**:** 85-92.

Montaut, A., Calvet, L., Bouvier, G., et al. (2013). "L'appariement handicap-santé et données de l'assurance maladie." Serie Sources Et Methodes - Document De Travail - Drees(39): 35. http://www.drees.sante.gouv.fr/serie-sources-et-methodes,755.html

Perlbarg, J., Allonier, C., Boisnault, P., et al. (2014). "[Feasibility and practical value of statistical matching of a general practice database and a health insurance database applied to diabetes and hypertension]." Sante Publique **26**(3): 355-363.

Toulmonde, M., Ducimetiere, F., Jouhet, V., et al. (2016). "Les grandes bases de donnees nationales et les etudes dans la vraie vie : l'exemple des sarcomes en France." Bull Cancer **103**(6 Suppl 1): S71-75.

Tuppin, P., Rudant, J., Constantinou, P., et al. (2017). "Value of a national administrative database to guide public decisions: From the systeme national d'information interregimes de l'Assurance Maladie (SNIIRAM) to the systeme national des donnees de sante (SNDS) in France." Rev Epidemiol Sante Publique **65 Suppl 4**: S149-s167.

Zins, M., Goldberg, M. et et al. (2018). L'appariement d'une cohorte épidémiologique aux bases de données médico-adminisratives. Santé et intelligence artificielle., Paris : CNRS éditions**:** 59-74.

Études étrangères

Quelques études clefs

Ben-Shalom, Y. et Stapleton, D. C. (2016). "Predicting Disability among Community-Dwelling Medicare Beneficiaries Using Claims-Based Indicators." Health Serv Res **51**(1): 262-281.

Iezzoni, L. I. (2002). "Using administrative data to study persons with disabilities." Milbank Q **80**(2): 347-379.

Palsbo, S. E., Sutton, C. D., Mastal, M. F., et al. (2008). "Identifying and classifying people with disabilities using claims data: further development of the Access Risk Classification System (ARCS) algorithm." Disabil Health J **1**(4): 215-223.

BASES DE L'ASSURANCE MALADIE

Allen, L. A., Yood, M. U., Wagner, E. H., et al. (2014). "Performance of claims-based algorithms for identifying heart failure and cardiomyopathy among patients diagnosed with breast cancer." Med Care **52**(5): e30-38.

> BACKGROUND: Cardiotoxicity is a known complication of certain breast cancer therapies, but rates come from clinical trials with design features that limit external validity. The ability to accurately identify cardiotoxicity from administrative data would enhance safety information. OBJECTIVE: To characterize the performance of clinical coding algorithms for identification of cardiac dysfunction in a cancer population. RESEARCH DESIGN: We sampled 400 charts among 6460 women diagnosed with incident breast cancer, tumor size >/= 2 cm or node positivity, treated within 8 US health care systems between 1999 and 2007. We abstracted medical records for clinical diagnoses of heart failure (HF) and cardiomyopathy (CM) or evidence of reduced left ventricular ejection fraction. We then assessed the performance of 3 different International Classification of Diseases, 9th Edition (ICD-9)-based algorithms. RESULTS: The HF/CM coding algorithm designed a priori to balance performance characteristics provided a sensitivity of 62% (95% confidence interval, 40%-80%), specificity of 99% (range, 97% to 99%), positive predictive value (PPV) of 69% (range, 45% to 85%), and negative predictive value (NPV) of 98% (range, 96% to 99%). When applied only to incident HF/CM (ICD-9 codes and gold standard diagnosis both occurring after breast cancer diagnosis) in patients exposed to anthracycline and/or trastuzumab therapy, the PPV was 42% (range, 14% to 76%). CONCLUSIONS: Claims-based algorithms have moderate sensitivity and high specificity for identifying HF/CM among patients with invasive breast cancer. As the prevalence of HF/CM among the breast cancer population is low, ICD-9 codes have high NPV but only moderate PPV. These findings suggest a significant degree of misclassification due to HF/CM overcoding versus incomplete clinical documentation of HF/CM in the medical record.

Ben-Shalom, Y. et Stapleton, D. C. (2016). "Predicting Disability among Community-Dwelling Medicare Beneficiaries Using Claims-Based Indicators." Health Serv Res **51**(1): 262-281.

> OBJECTIVES: To assess the feasibility of using existing claims-based algorithms to identify community-dwelling Medicare beneficiaries with disability based solely on the conditions for which they are being treated, and improving on these algorithms by combining them in predictive models. DATA SOURCE: Data on 12,415 community-dwelling fee-for-service Medicare beneficiaries who first responded to the Medicare Current Beneficiary Survey (MCBS) in 2003-2006. STUDY DESIGN: Logistic regression models in which six claims-based disability indicators are used to predict self-reported disability. Receiver operating characteristic (ROC) curves were used to assess the performance of the predictive models. PRINCIPAL FINDINGS: The predictive performance of the regression-based models is better than that of the individual claims-based indicators. At a predicted probability threshold chosen to maximize the

sum of sensitivity and specificity, sensitivity is 0.72 for beneficiaries age 65 or older and specificity is 0.65. For those under 65, sensitivity is 0.54 and specificity is 0.67. The findings also suggest ways to improve predictive performance for specific disability populations of interest to researchers. CONCLUSIONS: Predictive models that incorporate multiple claims-based indicators provide an improved tool for researchers seeking to identify people with disabilities in claims data.

Berger, A., Edelsberg, J., Inglese, G., et al. (2009). "Identification of patients receiving peritoneal dialysis using health insurance claims data." Clin Ther **31**(6): 1321-1334.

OBJECTIVE: The aim of this analysis was to assess alternative methods of identification of patients treated with peritoneal dialysis (PD) in health care claims databases for possible use in future analyses of costs of this treatment modality. METHODS: Using a US health insurance claims database spanning January 1, 2004, to December 31, 2006, we identified all patients with renal failure who satisfied a case-finding algorithm for PD anticipated to be highly specific, but not necessarily sensitive-namely, > or =2 claims for PD-related physician services (algorithm 1). All claims from these patients were assessed to identify additional PD-related codes, from which 6 additional algorithms were developed, each of which focused on specific categories of billing codes (eg, diagnostic, procedural/service, equipment). Patient selection was then reimplemented using these alternative algorithms. Concordance between the various algorithms and the extent to which resulting samples were similar in terms of patient characteristics, health care resource utilization, and costs were assessed. RESULTS: We identified a total of 132,274 patients in the database with > or =1 claim for renal failure and valid enrollment data. Among these patients, a total of 2329 satisfied case-selection criteria for algorithm 1, and 4031 patients met criteria for at least 1 of the 7 algorithms for PD. The most sensitive algorithm identified 2859 patients who might have received PD; the least sensitive, 211. Concordance between algorithms was relatively poor. Patients identified using each algorithm were similar, however, with respect to mean age (45-50 years), sex (54%-56% male), and the prevalence of selected comorbidities. Annualized median health care costs were similar across the various algorithms (range, US $80,967-$118,668). CONCLUSIONS: Based on the results from this analysis, it seems that health care providers bill insurers for PD-related care using a variety of codes. Investigators using health insurance claims data for analyses of patients treated with PD need to take this into account.

Blais, C. et Rochette, L. (2015). "Trends in prevalence, incidence and mortality of diagnosed and silent coronary heart disease in Quebec." Health Promot Chronic Dis Prev Can **35**(10): 184-193.

INTRODUCTION: Of all cardiovascular causes of mortality, coronary heart disease (CHD) remains the leading cause of death. Our objectives were to establish trends in the prevalence and incidence of CHD in the province of Quebec, and to determine the proportion of CHD mortality that had no previous CHD diagnosis. METHODS: Trends in prevalence, incidence and mortality were examined with a population-based study using the Quebec Integrated Chronic Disease Surveillance System, which links several health administrative databases. Data are presented using two case definitions for Quebecers aged 20 years and over: 1) a validated definition, and 2) CHD causes of death codes added to estimate the proportion of deaths that occurred without any previous CHD diagnosis as a proxy for sudden cardiac death (SCD). RESULTS: In 2012/2013, the crude prevalence of CHD was 9.4% with the first definition (593 000 people). Between 2000/2001 and 2012/2013, the age-standardized prevalence increased by 14%, although it has been decreasing slightly since 2009/2010. Age-standardized incidence and mortality rates decreased by 46% and 26% respectively, and represented a crude rate of 6.9 per 1000 and 5.2% in 2012/2013. The proportion identified only by CHD mortality, our SCD proxy, was only significant for the incident cases (0.38 per 1000 in 2009/2010) and declined over the study period. CONCLUSION: The prevalence of CHD has tended to decrease in recent years, and incidence and mortality have been declining in Quebec. Most CHD mortality occurs in previously diagnosed patients and only a small proportion of incident cases were not previously identified.

Chi, W. C., Chang, K. H., Escorpizo, R., et al. (2014). "Measuring disability and its predicting factors in a large database in Taiwan using the World Health Organization Disability Assessment Schedule 2.0." Int J Environ Res Public Health **11**(12): 12148-12161.

The definition of disability had been unclear until the International Classification of Functioning, Disability, and Health was promulgated in 2001 by the World Health Organization (WHO). Disability is a critical but relatively neglected public-health concern. We conducted this study to measure disabilities by using the WHO Disability Assessment Schedule 2.0 (WHODAS 2.0) and identify the factors that contribute to disabilities. We obtained and analyzed the data on people who applied to Taiwan's disability registration system between September 2012 and August 2013. A total of 158,174 cases were selected for this study. Among the people included in this study, 53% were male, and the females were on average 3 years older than the males. More males than females were of a low socioeconomic status, but the rate of employment was higher among the males than among the females. Age, sex, place of residence, and types and severity of impairment were all determined to be factors that independently contributed to disability. This study has demonstrated that disability can be measured and compared using WHODAS 2.0. Increasing the public-health attention devoted to disability and identifying the factors associated with disability can promote independence and social participation in people with disabilities.

Ciol, M. A., Rasch, E. K., Hoffman, J. M., et al. (2014). "Transitions in mobility, ADLs, and IADLs among working-age Medicare beneficiaries." Disabil Health J **7**(2): 206-215.

BACKGROUND: Disability is a dynamic process where functional status may change over time. Examination of the Medicare population suggests that, for those over age 65, disability status will fluctuate in 30% of beneficiaries each year. Less is known about those under age 65. The dynamic nature of disability is of relevance since it has important implications for social policies related to disability. OBJECTIVES: To: (1) describe the characteristics of Medicare beneficiaries eligible due to disability; and (2) estimate the proportion of individuals with transitions in functional status over a one-year period stratified by baseline characteristics and diagnostic subgroups. METHODS: We used the Medicare Current Beneficiary Survey from 1995 to 2005 to examine transitions in mobility and daily activities among individuals who were eligible for Medicare coverage due to disability. RESULTS: From the standpoint of function in mobility and daily activities, the working-age Medicare population with disability is fairly stable. While 75%-90% of our sample reported no disability or stable disability from one year to the next, depending on the condition and disability metric, as many as 13-14% of individuals showed improvement or decline in their functional status. CONCLUSIONS: In the working-age population with disability, a small percentage of individuals will improve or worsen from one year to the next. Since these transitions are associated with a variety of individual characteristics including health conditions, further research applied to larger samples is required to refine policy relevant models that might inform decisions related to ongoing eligibility for disability programs.

Davidoff, A. J., Gardner, L. D., Zuckerman, I. H., et al. (2014). "Validation of disability status, a claims-based measure of functional status for cancer treatment and outcomes studies." Med Care **52**(6): 500-510.

BACKGROUND: In prior research, we developed a claims-based prediction model for poor patient disability status (DS), a proxy measure for performance status, commonly used by oncologists to summarize patient functional status and assess ability of a patient to tolerate aggressive treatment. In this study, we implemented and validated the DS measure in 4 cohorts of cancer patients: early and advanced non-small cell lung cancers (NSCLC), stage IV estrogen receptor-negative (ER-) breast cancer, and myelodysplastic syndromes (MDS). DATA AND METHODS: SEER-Medicare data (1999-2007) for the 4 cohorts of cancer patients. Bivariate and multivariate logistic regression tested the association of the DS measure with designated cancer-directed treatments: early NSCLC (surgery), advanced NSCLC (chemotherapy), stage IV ER- breast cancer (chemotherapy), and MDS (erythropoiesis-stimulating agents). Treatment model fit was compared across model iterations. RESULTS: In both unadjusted and adjusted results, predicted poor DS was strongly associated with a lower likelihood of cancer treatment receipt in all 4 cohorts [early NSCLC (N=20,280), advanced NSCLC (N=31,341), stage IV ER- breast cancer (N=1519), and MDS (N=6058)] independent of other patient, contextual, and disease characteristics, as well as the Charlson Comorbidity Index. Inclusion of the DS measure into models already controlling for other variables did not significantly improve model fit across the cohorts.

CONCLUSIONS: The DS measure is a significant independent predictor of cancer-directed treatment. Small changes in model fit associated with both DS and the Charlson Comorbidity Index suggest that unobserved factors continue to play a role in determining cancer treatments.

Deshpande, A. D., Schootman, M. et Mayer, A. (2015). "Development of a claims-based algorithm to identify colorectal cancer recurrence." Ann Epidemiol **25**(4): 297-300.

PURPOSE: To examine the validity of claims data to identify colorectal cancer (CRC) recurrence and determine the extent to which misclassification of recurrence status affects estimates of its association with overall survival in a population-based administrative database. METHODS: We calculated the accuracy of claims data relative to medical records from one large tertiary hospital to identify CRC recurrence. We estimated the effect of misclassifying recurrence on survival by applying these findings to the linked Surveillance, Epidemiology, and End Results-Medicare data. RESULTS: Of 174 eligible CRC patients identified through medical records, 32 (18.4%) had a recurrence. A claims-based algorithm of secondary malignancy codes yielded a sensitivity of 81% and specificity of 99% for identifying recurrence. Agreement between data sources was almost perfect (kappa: 0.86). In a model unadjusted for misclassification, CRC patients with recurrence were 3.04 times (95% confidence interval: 2.92-3.17) more likely to die of any cause than those without recurrence. In the corrected model, CRC patients with recurrence were 3.47 times (95% confidence interval: 3.06-4.14) more likely to die than those without recurrence. CONCLUSIONS: Identifying recurrence in CRC patients using claims data is feasible with moderate sensitivity and high specificity. Future studies can use this algorithm with Surveillance, Epidemiology, and End Results-Medicare data to study treatment patterns and outcomes of CRC patients with recurrence.

Iezzoni, L. I. (2002). "Using administrative data to study persons with disabilities." Milbank Q **80**(2): 347-379.

Administrative data result from administering health plans--tracking service utilization, paying claims, monitoring costs and quality--and have been used extensively for health services research. This article examines the strengths and limitations of administrative data for health services research studies of people with disabilities. Administrative data offer important advantages: encompassing large populations over time, ready availability, low cost, and computer readability. Questions arise about how to identify people with disabilities, capture disability-related services, and determine meaningful health care outcomes. Potentially useful administrative data elements include eligibility for Medicare or Medicaid through Social Security disability determinations, diagnosis and procedure codes, pharmacy claims, and durable medical equipment claims. Linking administrative data to survey or other data sources enhances the utility of administrative data for disability studies.

Khoury, A. J., Hall, A., Andresen, E., et al. (2013). "The association between chronic disease and physical disability among female Medicaid beneficiaries 18-64 years of age." Disabil Health J **6**(2): 141-148.

BACKGROUND: Rates of physical disability are higher in women than in men, and economically disadvantaged women are at greater risk for physical disability than women with higher incomes. Chronic diseases increase the risk of physical disability, and people with physical disability experience some added risks of secondary conditions including chronic disease. Yet, little is known about the prevalence of chronic disease among women living with a physical disability who use Medicaid, a particularly disadvantaged population. OBJECTIVE: This study described the prevalence of chronic disease among adult (18-64 years), female, Florida Medicaid beneficiaries living with a physical disability between 2001 and 2005. METHODS: Using Medicaid eligibility and claims files, we extracted ICD-9 codes for physically-disabling conditions and Current Procedure Terminology codes for mobility-assistive devices to define three levels of physical disability. RESULTS: Participants appeared to be at high risk for both physical disability and chronic diseases. Close to half of the women had been diagnosed with one or more physically-disabling conditions, and 5.3% used mobility devices. One-third of the women had hypertension and sizeable proportions had other chronic diseases. Women with physical disability were more likely to have co-morbid chronic diseases than their able-bodied counterparts. DISCUSSION: Our findings support the need for improved chronic disease prevention

among female Medicaid beneficiaries, particularly those with physical disability. Strategies to improve prevention, screening and treatment in this population may mitigate the trends toward higher physical disability rates in the low-income, working-age population and may prevent high Medicare and Medicaid costs in the long-run.

Kohn, C. G., Weeda, E. R., Kumar, N., et al. (2017). "External validation of a claims-based and clinical approach for predicting post-pulmonary embolism outcomes among United States veterans." Intern Emerg Med **12**(5): 613-619.

The In-hospital Mortality for PulmonAry embolism using Claims daTa (IMPACT) rule can accurately identify pulmonary embolism (PE) patients at low risk of early complications using claims data. We sought to externally validate the IMPACT and simplified Pulmonary Embolism Severity Index (sPESI) tools for predicting all-cause mortality and readmission. We used Veteran Health Administration data (10/1/2010-9/30/2015) to identify adults with >/=1 inpatient diagnosis code for acute PE, >/=12 months continuous medical and pharmacy benefits prior to the index PE, >/=90 days of post-event follow-up (unless death occurred) and >/=1 claim for an anticoagulant during the index PE stay. Prognostic accuracies of IMPACT and sPESI for 30- and 90-day all-cause mortality and 90-day readmission were estimated. Of 6,746 PE patients, 7.5 and 12.6% died at 30 and 90 days. Within 90 days, 20.1% were readmitted for any reason. Hospitalization for recurrent VTE and major bleeding occurred in 5.6 and 1.7% of patients. IMPACT classified 15.2% as low risk, while 28.4% were low risk per sPESI. Both tools displayed sensitivity >90% and negative predictive values (NPVs) >97% for 30-day mortality, but low specificity (range 16.2-30.0) and positive predictive values (PPVs) (range 8.7-9.5); with similar results observed for 90-day mortality. IMPACT's sensitivity for all-cause readmission was numerically higher than sPESI (88.2 vs. 79.0%), but both had comparable NPVs (85.1 vs. 84.2%). Similar trends were observed for VTE or major bleeding readmissions. IMPACT classified patients for post-PE outcomes with similar accuracy as sPESI. IMPACT appears useful for identifying PE patients at low risk for early mortality or readmission in claims-based studies.

Kucharska-Newton, A. M., Heiss, G., Ni, H., et al. (2016). "Identification of Heart Failure Events in Medicare Claims: The Atherosclerosis Risk in Communities (ARIC) Study." J Card Fail **22**(1): 48-55.

BACKGROUND: We examined the accuracy of Medicare heart failure (HF) diagnostic codes in the identification of acute decompensated (ADHF and chronic stable (CSHF) HF. METHODS AND RESULTS: Hospitalizations were identified from medical discharge records for Atherosclerosis Risk in Communities (ARIC) study participants with linked Medicare Provider Analysis and Review (MedPAR) files for the years 2005-2009. The ARIC study classification of ADHF and CSHF, based on adjudicated review of medical records, was considered to be the criterion standard. A total 8,239 ARIC medical records and MedPAR records meeting fee-for-service (FFS) criteria matched on unique participant ID and date of discharge (68.5% match). Agreement between HF diagnostic codes from the 2 data sources found in the matched records for codes in any position (kappa > 0.9) was attenuated for primary diagnostic codes (kappa < 0.8). Sensitivity of HF diagnostic codes found in Medicare claims in the identification of ADHF and CSHF was low, especially for the primary diagnostic codes. CONCLUSION: Matching of hospitalizations from Medicare claims with those obtained from abstracted medical records is incomplete, even for hospitalizations meeting FFS criteria. Within matched records, HF diagnostic codes from Medicare show excellent agreement with HF diagnostic codes obtained from medical record abstraction. The Medicare data may, however, overestimate the occurrence of hospitalized ADHF or CSHF.

Lin, E., Balogh, R., Cobigo, V., et al. (2013). "Using administrative health data to identify individuals with intellectual and developmental disabilities: a comparison of algorithms." J Intellect Disabil Res **57**(5): 462-477.

BACKGROUND: Individuals with intellectual and developmental disabilities (IDD) experience high rates of physical and mental health problems; yet their health care is often inadequate. Information about their characteristics and health services needs is critical for planning efficient and equitable services. A logical source of such information is administrative health data; however, it can be difficult to identify

cases with IDD in these data. The purpose of this study is to evaluate three algorithms for case finding of IDD in health administrative data. METHODS: The three algorithms were created following existing approaches in the literature which ranged between maximising sensitivity versus balancing sensitivity and specificity. The broad algorithm required only one IDD service contact across all available data and time periods, the intermediate algorithm added the restriction of a minimum of two physician visits while the narrow algorithm added a further restriction that the time period be limited to 2006 onward. The resulting three cohorts were compared according to socio-demographic and clinical characteristics. Comparisons on different subgroups for a hypothetical population of 50,000 individuals with IDD were also carried out: this information may be relevant for planning specialised treatment or support programmes. RESULTS: The prevalence rates of IDD per 100 were 0.80, 0.52 and 0.18 for the broad, intermediate and narrow algorithms, respectively. Except for 'percentage with psychiatric co-morbidity', the three cohorts had similar characteristics (standardised differences < 0.1). More stringent thresholds increased the percentage of psychiatric co-morbidity and decreased the percentages of women and urban residents in the identified cohorts (standardised differences = 0.12 to 0.46). More concretely, using the narrow algorithm to indirectly estimate the number of individuals with IDD, a practice not uncommon in planning and policy development, classified nearly 7000 more individuals with psychiatric co-morbidities than using the intermediate algorithm. CONCLUSIONS: The prevalence rate produced by the intermediate algorithm most closely approximated the reported literature rate suggesting the value of imposing a two-physician visit minimum but not restricting the time period covered. While the statistical differences among the algorithms were generally minor, differences in the numbers of individuals in specific population subgroups may be important particularly if they have specific service needs. Health administrative data can be useful for broad-based service planning for individuals with IDD and for population level comparisons around their access and quality of care.

Palsbo, S. E., Sutton, C. D., Mastal, M. F., et al. (2008). "Identifying and classifying people with disabilities using claims data: further development of the Access Risk Classification System (ARCS) algorithm." <u>Disabil Health J</u> **1**(4): 215-223.

BACKGROUND: The goal was to develop an inexpensive and rapid method for health systems to classify people by their ability to access routine care. We sought to refine and revalidate a software algorithm, the Access Risk Classification System (ARCS), using automated claims data to classify people into one of four categories based on the probable need for care coordination or health system accommodations. METHODS: Through simple linkages of longitudinal claims data, the algorithm assigned individuals into one of four categories. We evaluated the algorithm's sensitivity and specificity by comparing the predicted classification against self-report. The validation results were used to refine the algorithm. RESULTS: When we classified people into two groups of any degree of functional limitation or no limitation, the sensitivity was 91% and the specificity was 26%. When classified into two groups of those needing proactive care coordination and all others, sensitivity was 83% and specificity was 30%. Thus, overall correct classification ranges from good to fair. CONCLUSIONS: The algorithm utilizes claims databases readily available to many health claims payers. Adding Healthcare Common Procedural Coding System claims and number of prescriptions improves correct classification rates. Even when the claims data are incomplete and imprecise, ARCSv2 (ARCS version 2) can be used as an initial screen to identify people who should be included in the calculation of quality measures and who should be surveyed for consumer reported quality measurement. When using four classification categories, 69% of the people with the greatest risk and need for care coordination are correctly identified. ARCS can increase the correct identification of people with disabilities by 400% over random digit dialing of a general population. However, the ARCS should be further refined and validated in a larger population that includes more men with disabilities, children, and people without disabilities before it is used to compute quality measures using administrative data. Correct classification might be improved by incorporating information on comorbidities and specific medication categories.

Saczynski, J. S., Andrade, S. E., Harrold, L. R., et al. (2012). "A systematic review of validated methods for identifying heart failure using administrative data." <u>Pharmacoepidemiol Drug Saf</u> **21 Suppl 1**: 129-140.

PURPOSE: To identify and describe the validity of algorithms used to detect heart failure (HF) using administrative and claims data sources. METHODS: A systematic review of PubMed and Iowa Drug Information Service searches of the English language was performed to identify studies published between 1990 and 2010 that evaluated the validity of algorithms for the identification of patients with HF using and claims data. Abstracts and articles were reviewed by two study investigators to determine their relevance on the basis of predetermined criteria. RESULTS: The initial search strategy identified 887 abstracts. Of these, 499 full articles were reviewed and 35 studies included data to evaluate the validity of identifying patients with HF. Positive predictive values (PPVs) were in the acceptable to high range, with most being very high (>90%). Studies that included patients with a primary hospital discharge diagnosis of International Classification of Diseases, Ninth Revision, code 428.X had the highest PPV and specificity for HF. PPVs for this algorithm ranged from 84% to 100%. This algorithm, however, may compromise sensitivity because many HF patients are managed on an outpatient basis. The most common 'gold standard' for the validation of HF was the Framingham Heart Study criteria. CONCLUSIONS: The algorithms and definitions used to identify HF using administrative and claims data perform well, particularly when using a primary hospital discharge diagnosis. Attention should be paid to whether patients who are managed on an outpatient basis are included in the study sample. Including outpatient codes in the described algorithms would increase the sensitivity for identifying new cases of HF.

Searles Nielsen, S., Warden, M. N., Camacho-Soto, A., et al. (2017). "A predictive model to identify Parkinson disease from administrative claims data." Neurology **89**(14): 1448-1456.

OBJECTIVE: To use administrative medical claims data to identify patients with incident Parkinson disease (PD) prior to diagnosis. METHODS: Using a population-based case-control study of incident PD in 2009 among Medicare beneficiaries aged 66-90 years (89,790 cases, 118,095 controls) and the elastic net algorithm, we developed a cross-validated model for predicting PD using only demographic data and 2004-2009 Medicare claims data. We then compared this model to more basic models containing only demographic data and diagnosis codes for constipation, taste/smell disturbance, and REM sleep behavior disorder, using each model's receiver operator characteristic area under the curve (AUC). RESULTS: We observed all established associations between PD and age, sex, race/ethnicity, tobacco smoking, and the above medical conditions. A model with those predictors had an AUC of only 0.670 (95% confidence interval [CI] 0.668-0.673). In contrast, the AUC for a predictive model with 536 diagnosis and procedure codes was 0.857 (95% CI 0.855-0.859). At the optimal cut point, sensitivity was 73.5% and specificity was 83.2%. CONCLUSIONS: Using only demographic data and selected diagnosis and procedure codes readily available in administrative claims data, it is possible to identify individuals with a high probability of eventually being diagnosed with PD.

Stein, J. D., Blachley, T. S. et Musch, D. C. (2013). "Identification of persons with incident ocular diseases using health care claims databases." Am J Ophthalmol **156**(6): 1169-1175.e1163.

PURPOSE: To assess the extent to which incidence rates calculated for common ocular diseases by using claims data may be overestimated according to the length of the disease-free look-back period used in the analysis. DESIGN: Retrospective longitudinal cohort analysis. METHODS: Billing records of 2457 persons continuously enrolled for 11 years in a managed-care network were searched for International Classification of Diseases (ICD-9-CM) diagnoses of cataract, open-angle glaucoma (OAG), nonexudative age-related macular degeneration (ARMD), and nonproliferative diabetic retinopathy (NPDR) at eye-care visits in the first half of 2001, the second half of 2010, and 2011. For each condition, incidence rates calculated by using "look-back" periods ranging from 0.5-9 years were compared with best estimates from a gold-standard period of 9.5 years. RESULTS: With a 1-year disease-free look-back period, incidence was overestimated by 260% for cataract, 135% for OAG, 209% for ARMD, and 300% for NPDR. Expanding the disease-free look-back period to 3 years resulted in a reduction of incidence overestimation to 40% for cataract, 14% for OAG, 45% for ARMD, and 100% for NPDR. A 5-year look-back period yielded incidence rates that were overestimated by <30% for all 4 conditions. CONCLUSIONS: In our claims-data analysis of 4 common ocular conditions, a

disease-free interval </=1 year insufficiently distinguished newly diagnosed from pre-existing disease, resulting in grossly overestimated incidence rates. Using look-back periods of 3-5 years, depending on the specific diagnosis, yielded considerably more accurate estimates of disease incidence.

Yasaitis, L. C., Berkman, L. F. et Chandra, A. (2015). "Comparison of self-reported and Medicare claims-identified acute myocardial infarction." Circulation **131**(17): 1477-1485; discussion 1485.

BACKGROUND: Cardiovascular disease is often studied through patient self-report and administrative data. However, these 2 sources provide different information, and few studies have compared them. METHODS AND RESULTS: We compared data from a longitudinal, nationally representative survey of older Americans with matched Medicare claims. Self-reported heart attack in the previous 2 years was compared with claims-identified acute myocardial infarction (AMI) and acute coronary syndrome. Among the 3.1% of respondents with self-reported heart attack, 32.8% had claims-identified AMI, 16.5% had non-AMI acute coronary syndrome, and 25.8% had other cardiac claims; 17.3% had no inpatient visits in the previous 2.5 years. Claims-identified AMIs were found in 1.4% of respondents; of these, 67.8% reported a heart attack. Self-reports were less likely among respondents >75 years of age (62.7% versus 74.6%; P=0.006), with less than high school education (61.6% versus 71.4%; P=0.015), with at least 1 limitation in activities of daily living (59.6% versus 74.7%; P=0.001), or below the 25th percentile of a word recall memory test (60.7% versus 71.3%; P=0.019). Both self-reported and claims-identified cardiac events were associated with increased mortality; the highest mortality was observed among those with claims-identified AMI who did not self-report (odds ratio, 2.8; 95% confidence interval, 1.5-5.1) and among those with self-reported heart attack and claims-identified AMI (odds ratio, 2.5; 95% confidence interval, 1.7-3.6) or non-AMI acute coronary syndrome (odds ratio, 2.7; 95% confidence interval, 1.8-4.1). CONCLUSIONS: There is considerable disagreement between self-reported and claims-identified events. Although self-reported heart attack may be inaccurate, it indicates increased risk of death, regardless of whether the self-report is confirmed by Medicare claims.

AUTRES BASES ADMINISTRATIVES

Bogardus, S. T., Jr., Towle, V., Williams, C. S., et al. (2001). "What does the medical record reveal about functional status? A comparison of medical record and interview data." Journal of general internal medicine **16**(11): 728-736.
https://www.ncbi.nlm.nih.gov/pubmed/11722685
https://www.ncbi.nlm.nih.gov/pmc/PMC1495285/

OBJECTIVE: Functional status measures are potent independent predictors of hospital outcomes and mortality. The study objective was to compare medical record with interview data for functional status. SUBJECTS AND METHODS: Subjects were 525 medical patients, aged 70 years or older, hospitalized at an academic medical center. Patient interviews determined status for 7 basic activities of daily living (BADLs) and 7 instrumental activities of daily living (IADLs). Medical records were reviewed to assess documentation of BADLs and IADLs. RESULTS: Most medical records contained no documentation of individual BADLs and IADLs (61% to 98% of records lacking documentation), with the exception of walking (24% of medical records lacking documentation). Impairment prevalence was lower in medical records than at interview for all BADLs and IADLs, and agreement between interview and medical record was poor (kappa < 0.40 for individual BADLs and IADLs). Sensitivity of the medical record for BADL and IADL impairment was poor (range 95% to 44%), using the interview as a reference standard. Sensitivity and specificity of the medical record for detection of BADL and IADL impairment changed substantially when records with nondocumentation of functional status were excluded or were assumed to be equivalent to independence. CONCLUSIONS: The results suggest that the medical record is a poor source of data on many functional status measures, and that assuming that nondocumentation of functional status is equivalent to independence may be unwarranted. Given the prognostic importance of functional status measures, the results highlight the importance of

developing reliable and efficient means of obtaining functional status information on hospitalized older patients.

Draper, W. R., Hawley, C. E., McMahon, B. T., et al. (2014). "ADA perceived disability claims: a decision-tree analysis." J Occup Rehabil **24**(2): 213-219.

INTRODUCTION: The purpose of this study is to examine the possible interactions of predictor variables pertaining to perceived disability claims contained in a large governmental database. Specifically, it is a retrospective analysis of US Equal Employment Opportunity Commission (EEOC) data for the entire population of workplace discrimination claims based on the "regarded as disabled" prong of the Americans with Disabilities Act (ADA) definition of disability. METHODS: The study utilized records extracted from a "master database" of over two million charges of workplace discrimination in the Integrated Mission System of the EEOC. This database includes all ADA-related discrimination allegations filed from July 26, 1992 through December 31, 2008. Chi squared automatic interaction detection (CHAID) was employed to analyze interaction effects of relevant variables, such as issue (grievance) and industry type. The research question addressed by CHAID is: What combination of factors are associated with merit outcomes for people making ADA EEOC allegations who are "regarded as" having disabilities? RESULTS: The CHAID analysis shows how merit outcome is predicted by the interaction of relevant variables. Issue was found to be the most prominent variable in determining merit outcome, followed by industry type, but the picture is made more complex by qualifications regarding age and race data. Although discharge was the most frequent grievance among charging parties in the perceived disability group, its merit outcome was significantly less than that for the leading factor of hiring.

Horner-Johnson, W., Dobbertin, K., Lee, J. C., et al. (2014). "Disparities in health care access and receipt of preventive services by disability type: analysis of the medical expenditure panel survey." Health Serv Res **49**(6): 1980-1999.

OBJECTIVE: To examine differences in access to health care and receipt of clinical preventive services by type of disability among working-age adults with disabilities. DATA SOURCE: Secondary analysis of Medical Expenditure Panel Survey (MEPS) data from 2002 to 2008. STUDY DESIGN: We conducted cross-sectional logistic regression analyses comparing people with different types of disabilities on health insurance status and type; presence of a usual source of health care; delayed or forgone care; and receipt of dental checkups and cancer screening. DATA COLLECTION: We pooled annualized MEPS data files across years. Our analytic sample consisted of adults (18-64 years) with physical, sensory, or cognitive disabilities and nonmissing data for all variables of interest. PRINCIPAL FINDINGS: Individuals with hearing impairment had better health care access and receipt than people with other disability types. People with multiple types of limitations were especially likely to have health care access problems and unmet health care needs. CONCLUSIONS: There are differences in health care access and receipt of preventive care depending on what type of disability people have. More in-depth research is needed to identify specific causes of these disparities and assess interventions to address health care barriers for particular disability groups.

McKay, K. A., Tremlett, H., Fisk, J. D., et al. (2018). "Psychiatric comorbidity is associated with disability progression in multiple sclerosis." Neurology **90**(15): e1316-e1323.

OBJECTIVE: Emerging evidence suggests that comorbidity may influence disability outcomes in multiple sclerosis (MS); we investigated the association between psychiatric comorbidity and MS disability progression in a large multiclinic population. METHODS: This retrospective cohort study accessed prospectively collected information from linked clinical and population-based health administrative databases in the Canadian provinces of British Columbia and Nova Scotia. Persons with MS who had depression, anxiety, or bipolar disorder were identified using validated algorithms using physician and hospital visits. Multivariable linear regression models fitted using an identity link with generalized estimating equations were used to determine the association between psychiatric comorbidity and disability using all available Expanded Disability Status Scale (EDSS) scores. RESULTS:

A total of 2,312 incident cases of adult-onset MS were followed for a mean of 10.5 years, during which time 35.8% met criteria for a mood or anxiety disorder. The presence of a mood or anxiety disorder was associated with a higher EDSS score (beta coefficient = 0.28, p = 0.0002, adjusted for disease duration and course, age, sex, socioeconomic status, physical comorbidity count, and disease-modifying therapy exposure). Findings were statistically significant among women (beta coefficient = 0.31, p = 0.0004), but not men (beta coefficient 0.22, p = 0.17). CONCLUSION: Presence of psychiatric comorbidities, which were common in our incident MS cohort, increased the severity of subsequent neurologic disability. Optimizing management of psychiatric comorbidities should be explored as a means of potentially mitigating disability progression in MS.

McKenzie, K., Martin, L. et Ouellette-Kuntz, H. (2017). "Needles in the haystack: Using open-text fields to identify persons with intellectual and developmental disabilities in administrative home care data." Res Dev Disabil **69**: 85-95.

BACKGROUND: Use of administrative health data to study populations of interest is becoming more common. Identifying individuals with intellectual and developmental disabilities (IDD) in existing databases can be challenging due to inconsistent definitions and terminologies of IDD over time and across sectors, and the inability to rely on etiologies of IDD as they are frequently unknown. AIMS: To identify diagnoses related to IDD in an administrative database and create a cohort of persons with IDD. METHODS: Open-text diagnostic entries related to IDD were identified in an Ontario home care database (2003-2015) and coded as being either acceptable (e.g. Down syndrome) or ambiguous (e.g. intellectually challenged). The cognitive and functional skills of the resulting groups were compared using logistic regressions and standardized differences, and their age distributions were compared to that of the general home care population. RESULTS: Just under 1% of the home care population had a diagnostic entry related to IDD. Ambiguous terms were most commonly used (61%), and this group tended to be older and less impaired than the group with more acceptable terms used to describe their IDD. CONCLUSIONS: Open-text diagnostic variables in administrative health records can be used to identify and study individuals with IDD. IMPLICATIONS: Future work is needed to educate assessors on the importance of using standard, accepted terminology when recording diagnoses related to IDD.

Zhang, Q., Yang, Y. et Saver, J. L. (2015). "Discharge destination after acute hospitalization strongly predicts three month disability outcome in ischemic stroke." Restor Neurol Neurosci **33**(5): 771-775.

PURPOSE: Final global disability 3 months post-stroke is the cardinal endpoint in acute stroke clinical trials. The most similar variable available in administrative datasets is discharge destination at end of the acute hospitalization. We investigated the predictive value of discharge destination for final global disability. METHODS: In the public dataset of the two NINDS-TPA trials, we characterized discharge destination as a 4 level ordinal variable. Correlation coefficients and logistic models probed the relation with the modified Rankin Scale (mRS) of global disability at 3 months. RESULTS: Among the 624 ischemic stroke patients, discharge destination was home in 42.1% , rehabilitation in 33.0% , skilled nursing facility in 9.8% , and in hospital-death in 12.4% . A strong correlation was noted between hospital discharge destination and 3 month mRS, r = 0.71, P < 0.001. Length of stay showed a weaker correlation with 3 month mRS, r = 0.29, p < 0.0001. A multiple logistic regression model identified 4 categories of independent predictors of 3 month global disability outcome, with discharge destination as the strongest independent variable. CONCLUSIONS: Discharge destination is a powerful predictor of final 3 month global disability outcomes and a valid outcome measure for use in local and national quality improvement programs.

Zhang, T., Tremlett, H., Zhu, F., et al. (2018). "Effects of physical comorbidities on disability progression in multiple sclerosis." Neurology **90**(5): e419-e427.

OBJECTIVE: To examine the association between physical comorbidities and disability progression in multiple sclerosis (MS). METHODS: We conducted a retrospective cohort study using linked health administrative and clinical databases in 2 Canadian provinces. Participants included adults with incident MS between 1990 and 2010 who entered the cohort at their MS symptom onset date.

Comorbidity status was identified with validated algorithms for health administrative data and was measured during the 1 year before study entry and throughout the study period. The outcome was the Expanded Disability Status Scale (EDSS) score as recorded at each clinic visit. We used generalized estimating equations to examine the association between physical comorbidities and EDSS scores over time, adjusting for sex, age, cohort entry year, use of disease-modifying drugs, disease course, and socioeconomic status. Meta-analyses were used to estimate overall effects across the 2 provinces. RESULTS: We identified 3,166 individuals with incident MS. Physical comorbidity was associated with disability; with each additional comorbidity, there was a mean increase in the EDSS score of 0.18 (95% confidence interval [CI] 0.09-0.28). Among specific comorbidities, the presence of ischemic heart disease (IHD) or epilepsy was associated with higher EDSS scores (IHD 0.31, 95% CI 0.01-0.61; epilepsy 0.68, 95% CI 0.11-1.26). CONCLUSIONS: Physical comorbidities are associated with an apparent increase in MS disability progression. Appropriate management of comorbidities needs to be determined to optimize outcomes.

METHODES, MODELES

Hammill, B. G., Hernandez, A. F., Peterson, E. D., Fonarow, G. C., Schulman, K. A. et Curtis, L. H. (2009). "Linking inpatient clinical registry data to Medicare claims data using indirect identifiers." American heart journal **157**(6): 995-1000.
https://www.ncbi.nlm.nih.gov/pubmed/19464409
https://www.ncbi.nlm.nih.gov/pmc/PMC2732025/

BACKGROUND: Inpatient clinical registries generally have limited ability to provide a longitudinal perspective on care beyond the acute episode. We present a method to link hospitalization records from registries with Medicare inpatient claims data, without using direct identifiers, to create a unique data source that pairs rich clinical data with long-term outcome data. METHODS AND RESULTS: The method takes advantage of the hospital clustering observed in each database by demonstrating that different combinations of indirect identifiers within hospitals yield a large proportion of unique patient records. This high level of uniqueness also allows linking without advance knowledge of the Medicare provider number of each registry hospital. We applied this method to 2 inpatient databases and were able to identify 81% of 39,178 records in a large clinical registry of patients with heart failure and 91% of 6,581 heart failure records from a hospital inpatient database. The quality of the link is high, and reasons for incomplete linkage are explored. Finally, we discuss the unique opportunities afforded by combining claims and clinical data for specific analyses. CONCLUSIONS: In the absence of direct identifiers, it is possible to create a high-quality link between inpatient clinical registry data and Medicare claims data. The method will allow researchers to use existing data to create a linked claims-clinical database that capitalizes on the strengths of both types of data sources.

Lopez, M. A., Benavides, F. G., Alonso, J., et al. (2014). "[The value of using administrative data in public health research: the Continuous Working Life Sample]." Gac Sanit **28**(4): 334-337.

The use of administrative data is common practice in public health research. The present field note describes the Continuous Working Life Sample (CWLS) and its use in health research. The CWLS is built on records generated by all contacts with the social security system (work contracts, disability, etc.), plus tax data (monetary gains, income, etc.) and census data (level of education, country of birth, etc.), but does not allow individuals to be identified. The CWLS was started in 2004 with 4% (1.1 million persons) of the total population who were either contributors to or beneficiaries of the social security system. The information on the individuals in the CWLS is updated annually and lost individuals are replaced. This continuous design allows the construction of a cohort with information on working life and financial status and evaluation of their relationship with work disability. Future connection with clinical records would enable analysis of other health-related outcomes.

Norouzi, J., Yadollahpour, A., Mirbagheri, S. A., et al. (2016). "Predicting Renal Failure Progression in Chronic Kidney Disease Using Integrated Intelligent Fuzzy Expert System." Comput Math Methods Med **2016**: 6080814.

BACKGROUND: Chronic kidney disease (CKD) is a covert disease. Accurate prediction of CKD progression over time is necessary for reducing its costs and mortality rates. The present study proposes an adaptive neurofuzzy inference system (ANFIS) for predicting the renal failure timeframe of CKD based on real clinical data. METHODS: This study used 10-year clinical records of newly diagnosed CKD patients. The threshold value of 15 cc/kg/min/1.73 m(2) of glomerular filtration rate (GFR) was used as the marker of renal failure. A Takagi-Sugeno type ANFIS model was used to predict GFR values. Variables of age, sex, weight, underlying diseases, diastolic blood pressure, creatinine, calcium, phosphorus, uric acid, and GFR were initially selected for the predicting model. RESULTS: Weight, diastolic blood pressure, diabetes mellitus as underlying disease, and current GFR(t) showed significant correlation with GFRs and were selected as the inputs of model. The comparisons of the predicted values with the real data showed that the ANFIS model could accurately estimate GFR variations in all sequential periods (Normalized Mean Absolute Error lower than 5%). CONCLUSIONS: Despite the high uncertainties of human body and dynamic nature of CKD progression, our model can accurately predict the GFR variations at long future periods.

Palsbo, S. E. (2005). "Creating a disability registry using claims data." Managed care **13**(1): 6-12.

Silenou, B. C., Avalos, M., Helmer, C., et al. (2019). "Health administrative data enrichment using cohort information: Comparative evaluation of methods by simulation and application to real data." PLoS One **14**(1): e0211118.

BACKGROUND: Studies using health administrative databases (HAD) may lead to biased results since information on potential confounders is often missing. Methods that integrate confounder data from cohort studies, such as multivariate imputation by chained equations (MICE) and two-stage calibration (TSC), aim to reduce confounding bias. We provide new insights into their behavior under different deviations from representativeness of the cohort. METHODS: We conducted an extensive simulation study to assess the performance of these two methods under different deviations from representativeness of the cohort. We illustrate these approaches by studying the association between benzodiazepine use and fractures in the elderly using the general sample of French health insurance beneficiaries (EGB) as main database and two French cohorts (Paquid and 3C) as validation samples. RESULTS: When the cohort was representative from the same population as the HAD, the two methods are unbiased. TSC was more efficient and faster but its variance could be slightly underestimated when confounders were non-Gaussian. If the cohort was a subsample of the HAD (internal validation) with the probability of the subject being included in the cohort depending on both exposure and outcome, MICE was unbiased while TSC was biased. The two methods appeared biased when the inclusion probability in the cohort depended on unobserved confounders. CONCLUSION: When choosing the most appropriate method, epidemiologists should consider the origin of the cohort (internal or external validation) as well as the (anticipated or observed) selection biases of the validation sample.