Digital Personalized Health and Medicine
L.B. Pape-Haugaard et al. (Eds.)

doi:10.3233/SHTI200339

Impact Analysis of the Policy for Access of Administrative Data in France : A Before-After Study

Vincent LOOTEN[a,1] and Marie SIMON[b]
[a] *UMRS 1138, Centre de Recherche des Cordeliers, Université de Paris, Paris, France*
[b] *Université Paris-Est Créteil, Créteil, France*

Abstract. In 2017, French institutions reformed their data access policy regarding the national insurance and administrative databases systems (French abbreviation: SNDS), including claims data from hospitalization discharge summaries for the entire population follow-up encompassing over 10 years. Our study aimed to quantitatively and qualitatively describe such authorization before and after the reform. We extracted data access demands for French National Health Data Institute (INDS) data before and after the reforms. We included only studies that needed data extracted from the SNDS database and authorization of the regulator. We inferred the number of projects accepted pre- and post-reform, and we describe the types of studies, their topics, and the types of data used. We included 802 data access demands between January 1st 2008 and September 21st 2019. The median of data access demands by year increased from 21.5 to 203. This increase was lower in the studies included insurance data (21.5 to 70). The evolution is driven by the activity of Private companies and contract-research organization. The number of studies on Hematology and oncology and internal medicine increased respectively by 1.7 and 1.4 factors. Data access of claims data refers to the "accessible" dimension of the FAIR guiding principles. However, extrinsic factors influence the accessibility of claims data such as human factors (e.g. data scientist with experience in claims data) and economic factors (e.g. data infrastructure HIPAA and GDPR compliant).

Keywords. Data Sharing, Policy, Administrative Claims

1. Introduction

Data sharing is a prerequisite for open innovation in health to produce reproducible research and to promote open science. Healthcare professionals and citizens need to be equipped to share data to facilitate scientific research and minimize the potential privacy threats[1].Since May 2018, data sharing, including that of European citizens' personal data, as well as data processing in the European Union, has been subject to the General Data Protection Regulation (GDPR)[2]. The GDPR provides a legal framework for data reuse and offers a chance to standardize data protection practices in research, as well as opportunities for researchers in medical informatics to develop new models for information technology infrastructure[3]. In 2017, French institutions reformed their data access policy regarding the national insurance and administrative databases systems

[1] Corresponding Author: Dr Vincent Looten, Université de Paris, Paris, France; E-mail: lootenv@gmail.com.

(French abbreviation: SNDS), including claims data from hospitalization discharge summaries for the entire population follow-up encompassing over 10 years. Researchers with access to the SNDS are producing studies on various topics such as patient care pathways, pharmacoeconomics, and pharmacoepidemiology[4]. The French reform for data access to SNDS aimed to open insurance data to a wider research community, including public and private research labs. The research community hypothesized that enforcement of the GDPR has impacted this reform's effects. However, to the best of our knowledge, no previous studies have described the 2017 reform's impact on data access authorization with regard to the influence of the GDPR. Our study aimed to quantitatively and qualitatively describe such authorization before and after the reform.

2. Method

Data sources. We extracted data access demands for French National Health Data Institute (INDS) data before and after the reforms. The data used in this study are available at www.github.com/vlooten/datapolicy, while more recent data can be downloaded from the INDS website (www.indsante.fr).

Inclusion criteria. We included only studies that needed data extracted from the SNDS database and authorization of the regulator. Projects that were resubmitted were counted only once (for the first submission).

Data access policy. Before the reform, a national committee evaluated submitted projects, and an approval from the French Data Protection Authority (CNIL) was required (i.e., two authorizations were required). After the reform, all projects following the CNIL national guidelines required just one authorization from a national committee. This committee checked concordance between objectives and methods, and compatibility of the study objectives with the public interest.

Outcomes and definitions. We inferred the number of projects accepted pre- (before August 1st 2018) and post-reform (after August 1st 2018), and we describe the type of investigator, the types of studies, their topics, and the types of data used. The type of investigators described were: (1) Government and agencies (2) Private companies and Contract-research organizations (3) Hospitals and care structures (4) Citizens and associations (5) Schools and research centers. The categories of the type of the studies were: (1) Descriptive and transversal studies (2) Longitudinal or prognosis studies (3) Pharmacoepidemiology, pharmacoeconomic studies (4) Economics studies (5) Other (included methodological and unclassified studies). The topics described were: (1) Cardiology and vascular medicine (2) Hematology and oncology (3) Internal medicine (4) Psychiatry and addictology (5) Public health and occupational medicine (6) Surgery (7) Primary cares and (8) Other (included unclassified and unknown studies). The types of data described were: National data with or without insurance data.

Statistical analyses. Data were expressed as number (%). Chi2 tests (for categorical data) was used to compare groups. All tests involved use of R 3.6.1(R Foundation, Vienna, Austria).

3. Results

We reviewed 2031 data access demands between January 1st 2008 and September 21st 2019. We included 802 data access demands according to the inclusion criteria. Reasons

of exclusions were: no pairing with national administrative data (N=1146) data access rejected or canceled studies (N=41), renew data access authorization (N=32) and missing data (N=10). Years of data access demands were: 2019 (N=203), 2018 (N=257), 2017 (N=74), 2016 (N=73), 2015 (N=61), 2014 (N=51), 2013 (N=24), 2012 (N=17), 2011 (N=19), 2010 (N=15), 2009(N=6) and 2008 (N=2). The median number of data access demands by year increase from 21.5 to 203. Table 1 presents the characteristics of the data access demands. We observed a difference between the two groups for the type of investigator. The number of access of private companies increases from 26 (8.87%) to 227 (44.6%), and school and research centers decreased from 98 (33.4%) to 59 (11.6%). We observed a change in the type of the studies with a reduction of Pharmacoepidemiology, pharmacoeconomic studies (from 100 (34.1%) to 99 (19.4%). The structure of the topics also changes. In particular, the number of Hematology and oncology studies increased from 27 (9.22%) to 81 (15.9%).

Table 1. Description of the studies before and after the reform

	Before the reform (N=293)	**After the reform (N=509)**	**p value**
Insurance data	293	343	<0.001
Investigator:	293 (100%)	343 (67.4%)	<0.001
Government and Agencies	12 (4.10%)	27 (5.30%)	
Private companies and CRO	26 (8.87%)	227 (44.6%)	
Hospitals and care structures	125 (42.7%)	170 (33.4%)	
Citizens and associations	32 (10.9%)	26 (5.11%)	
Schools and research centers	98 (33.4%)	59 (11.6%)	
Type of study:			<0.001
Descriptive and transversal studies	48 (16.4%)	91 (17.9%)	
Economics studies	22 (7.51%)	47 (9.23%)	
Longitudinal or prognosis studies	116 (39.6%)	254 (49.9%)	
Pharmacoepidemiology, pharmacoeconomic studies	100 (34.1%)	99 (19.4%)	
Unknown	7 (2.39%)	18 (3.54%)	
Topics:			<0.001
Cardiology and vascular medicine	29 (9.90%)	51 (10.0%)	
Hematology and oncology	27 (9.22%)	81 (15.9%)	
Internal medicine	69 (23.5%)	164 (32.2%)	
Public health and occupational medicine	66 (22.5%)	102 (20.0%)	
Psychiatry and addictology	36 (12.3%)	50 (9.82%)	
Surgery			
Primary cares	5 (0.98%)	12 (4.10%)	
Other	11 (3.75%)	20 (3.93%)	

Over the 802 studies included, only 636 required insurance data. In this subgroup, the median of data access demands by year increased from 21.5 to 70. Table 2 presents the characteristics of the data access demands of this subgroup. The difference observed between the two groups are similar.

Table 2. Description of the studies included insurance data before and after the reform.

	Before the reform (N=293)	After the reform (N=343)	p value
Investigator:			<0.001
Government and Agencies	12 (4.10%)	24 (7.00%)	
Private companies and CRO	26 (8.87%)	129 (37.6%)	
Hospitals and care structures	125 (42.7%)	124 (36.2%)	
Citizens and associations	32 (10.9%)	17 (4.96%)	
Schools and research centers	98 (33.4%)	49 (14.3%)	
Type of study:			0.003
Descriptive and transversal studies	48 (16.4%)	50 (14.6%)	
Economics studies	22 (7.51%)	38 (11.1%)	
Longitudinal or prognosis studies	116 (39.6%)	172 (50.1%)	
Pharmacoepidemiology, pharmacoeconomic studies	100 (34.1%)	73 (21.3%)	
Unknown	7 (2.39%)	10 (2.92%)	
Topics:			<0.001
Cardiology and vascular medicine	29 (9.90%)	31 (9.04%)	
Hematology and oncology	27 (9.22%)	56 (16.3%)	
Internal medicine	69 (23.5%)	122 (35.6%)	
Public health and occupational medicine	66 (22.5%)	61 (17.8%)	
Psychiatry and addictology	36 (12.3%)	34 (9.91%)	
Surgery	43 (14.7%)	25 (7.29%)	
Primary cares	12 (4.10%)	4 (1.17%)	
Other	11 (3.75%)	10 (2.92%)	

4. Discussion

Main results. To the best of our knowledge, we are the first to describe the effect of the French reform on the data access demands. After the reform, the median of data access demands by year increased from 21.5 to 203. This increase was lower in the studies included insurance data (21.5 to 70). The evolution is driven by the activity of private companies and contract-research organization. The number of studies on Hematology and oncology and internal medicine increased respectively by 1.7 and 1.4 factors.

Technical significance. According to the FAIR guiding principles[5] data must be Findable, Accessible, Interoperable and Reusable. Claims data have important intrinsic

characteristics in the FAIR perspective. Data access of claims data refers to the "accessible" dimension of the FAIR guiding principles. However, extrinsic factors influence the accessibility of claims data such as human factors (e.g. data scientist with experience in claims data) and economic factors (e.g. data infrastructure HIPAA and GDPR compliant). Two reasons could explain the most important demands of private operators: (1) the reform has reduced the time of treatment (only one committee compounded by specialists of data) (2) and the legal framework has been simplified ensuring a good visibility for elaborate business plans. However, the activity of the public sector has increased fewer; this could be explained by the absence of new data infrastructure and new funding strategy (extrinsic factors). In our perspective, the efficiency of a data access policy should be quantitatively and qualitatively monitored to understand the potential leverages of data reuse.

Remaining challenges. Schmidt *et al.*[6] have described the natural history of the data reuse in the Danish National Patient Registry. However, in the French context, data access demands database is not linked to scientific publications. A perspective of our work is linking data access demands and publication to describe the natural history of the French administrative data reuse. Data reuse of claim database has a natural history and this history is influenced by the data access policy.

References

[1] S. Patil, H. Lu, C.L. Saunders, D. Potoglou, and N. Robinson, Public preferences for electronic health data storage, access, and sharing — evidence from a pan-European survey, *J. Am. Med. Informatics Assoc.* **23** (2016) 1096–1106. doi:10.1093/jamia/ocw012.

[2] J. Demotes-Mainard, C. Cornu, A. Guérin, P.-H. Bertoye, R. Boidin, S. Bureau, J.-M. Chrétien, C. Delval, D. Deplanque, C. Dubray, L. Duchossoy, V. Edel, C. Fouret, A. Galaup, F. Lesaulnier, M. Matei, F. Naudet, V. Plattner, M. Rubio, F. Ruiz, S. Sénéchal-Cohen, T. Simon, A. Vidal, A. Viola, and M. Violleau, How the new European data protection regulation affects clinical research and recommendations?, *Therapies*. **74** (2019) 31–42. doi:10.1016/j.therap.2018.12.004.

[3] J.M.M. Rumbold, and B. Pierscionek, The Effect of the General Data Protection Regulation on Medical Research, *J. Med. Internet Res.* **19** (2017) e47. doi:10.2196/jmir.7108.

[4] P. Tuppin, J. Rudant, P. Constantinou, C. Gastaldi-Ménager, A. Rachas, L. de Roquefeuil, G. Maura, H. Caillol, A. Tajahmady, J. Coste, C. Gissot, A. Weill, and A. Fagot-Campagna, Value of a national administrative database to guide public decisions: From the système national d'information interrégimes de l'Assurance Maladie (SNIIRAM) to the système national des données de santé (SNDS) in France, Rev. Epidemiol. Sante Publique. (2017). doi:10.1016/j.respe.2017.05.004.

[5] M.D. Wilkinson, M. Dumontier, Ij.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W.Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O.Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble,J.S. Grethe, J. Heringa, P.A.. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A.Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T.Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J.Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, and B. Mons, The FAIR Guiding Principles for scientific data management and stewardship, *Sci. Data*. **3** (2016) 160018. doi:10.1038/sdata.2016.18.

[6] M. Schmidt, S.A.J. Schmidt, J.L. Sandegaard, V. Ehrenstein, L. Pedersen, and H.T. Sørensen, The Danish National Patient Registry: a review of content, data quality, and research potential, *Clin. Epidemiol.* (2015) 449. doi:10.2147/CLEP.S91125.