

Assessment of the Accuracy of Using *ICD-10* Codes to Identify Systemic Sclerosis

This article was published in the following Dove Press journal:
Clinical Epidemiology

Sébastien De Almeida Chaves¹
Hélène Derumeaux²
Phuong Do Minh¹
Maryse Lapeyre-Mestre³⁻⁵
Guillaume Moulis^{1,4,5}
Grégory Pugnet^{1,4,5}

¹Department of Internal Medicine, CHU Toulouse, Toulouse, France; ²Department of Medical Information, CHU Toulouse, Toulouse, France; ³Department of Medical and Clinical Pharmacology, CHU Toulouse, Toulouse, France; ⁴UMR 1027 Inserm-University of Toulouse, Toulouse, France; ⁵Clinical Investigation Center 1436, CHU Toulouse, Toulouse, France

Importance: With the increased use of data from electronic medical records for research, it is important to validate in-patient electronic health records/hospital electronic health records for specific diseases identification using International Classification of Diseases, Tenth Revision (*ICD-10*) codes.

Objective: To assess the accuracy of using *ICD-10* codes to identify systemic sclerosis (SSc) in the French hospital database.

Design, Setting, and Participants: Electronic health record database analysis. The setting of the study's in-patient database was the Toulouse University Hospital, a tertiary referral center (2880 beds) that serves approximately 2.9 million inhabitants. Participants were patients with *ICD-10* discharge diagnosis codes of SSc seen at Toulouse University Hospital between January 1, 2010, and December 31, 2017.

Main Outcomes and Measures: The main outcome was the positive predictive value (PPV) of discharge diagnosis codes for identifying SSc. The PPVs were calculated by determining the ratio of the confirmed cases found by medical record review to the total number of cases identified by *ICD-10* code.

Results: Of the 2766 hospital stays, 216 patients were identified by an SSc discharge diagnosis code. Two hundred were confirmed as SSc after medical record review. The overall PPV was 93% (95% CI, 88–95%). The PPV for limited cutaneous SSc was 95% (95% CI, 85–98%).

Conclusions and Relevance: Our results suggest that using *ICD-10* codes alone to capture SSc is reliable in The French hospital database.

Keywords: systemic sclerosis, International Classification of Diseases, positive predictive value, sensitivity, hospital database

Methods

We conducted a validation study using the PMSI (Programme de Médicalisation des Systèmes d'Informations) data, the in-patient health database of the French Health Insurance System Database called SNDS (Système national des données de santé, <https://www.snds.gouv.fr/>)⁶ of Toulouse University Hospital (South of France, 4th city by population in the country, 2880 beds in the hospital) and the Toulouse University Hospital Systemic Sclerosis cohort (n=419).

Studies assessing the accuracy of diagnosis coding by medical chart review are authorized by the National Commission of Information Technology and Liberty (CNIL: French Law on Privacy decision No. 89–117). All hospitalized patients in this hospital received written information on the possibility of such studies.

Correspondence: Grégory Pugnet
Service de Médecine Interne, CHU
Toulouse Purpan, 1 Place du Dr Joseph
Baylac, TSA 40031, 31059, Toulouse
Cedex 9, France
Tel +33 5 61 77 71 26
Fax +33 5 61 77 71 24
Email pugnet.g@chu-toulouse.fr

Data Source

In all private and public hospitals in France, data on all hospital stays are recorded in the PMSI since 2007. Data recorded are the hospital identifier, the date of start and of end of stays, the diagnosis codes, the procedures, the costly drugs dispensed, and whether the stays occurred in a special unit (e.g., resuscitation or intensive care).^{6,7} The diagnoses are encoded using the ICD-10⁸ by the physician in charge of the patient or by specialized technician that generates the codes from medical charts. For this study, we used Toulouse University Hospital PMSI data corresponding to all stays from January 1, 2010, to December 31, 2017, in this hospital. In this institution, discharge diagnoses are coded by a specialized technician according to national standards. The quality of coding is regularly audited.⁹

The Toulouse University Hospital Systemic Sclerosis cohort (n=419) includes all consecutive SSc hospital stays at Toulouse University Hospital since 2010 who met the ACR/EULAR 2013 classification criteria for SSc.¹⁰ In these database, patient information is recorded electronically.

Population

As previously described,¹¹ we extracted all hospital stays with a primary discharge diagnosis ICD-10 code related to SSc: M34.0: “Progressive systemic sclerosis”, M34.1: “CR(E)ST syndrome”, M34.8: “Other forms of systemic sclerosis”, M34.8 + G73.7: “Other forms of systemic sclerosis with lung involvement”, M38.4 + J99.1: “Other forms of systemic sclerosis with myopathy” and M34.9: “Systemic sclerosis, unspecified” from January 1, 2010, to December 31.

Medical Chart Review

All medical records from this electronic query were individually and independently reviewed by two physicians specialized in SSc (S.D.A. and G.P.) to adjudicate the diagnosis. In the event of a discrepancy, a consensus was reached by discussion (it occurred for two patients). After reviewing medical records, cases were categorized as confirmed (true positive) if the ICD-10 code listed in the administrative database correctly identified the disease of interest (i.e., met the ACR/EULAR 2013 classification criteria for SSc).¹⁰ If the ICD-10 code did not match, it was noted as miscoded (false positive), and the actual diagnosis was determined.

Measured Outcomes and Statistical Analysis

The positive predictive value (PPV) and 95% confidence interval (CI) of ICD-10 codes were calculated as the ratio of the confirmed cases found by medical record review to the total number of cases identified by ICD-10 code.

For assessing the sensitivity, we identified all patients included in the Toulouse University Hospital Systemic Sclerosis cohort, who met the ACR/EULAR 2013 classification criteria for SSc,¹⁰ between 2010 and 2017 and who were admitted at Toulouse University Hospital. We then assessed whether their hospital stays were encoded using an ICD-10 code corresponding to an SSc disease. All analyses were performed using statistical software SAS[®], version 9.4.

Results

Of the 2766 hospital stays during the study period, 216 patients had an SSc discharge diagnosis code. Mean (SD) age was 62 (14) years and 163 patients (75.5%) were females. Medical record review confirmed 200 SSc cases (Table 1).

The overall PPV of SSc-specific ICD-10 codes to identify patients with SSc was 93% (95% CI, 88–95%). The PPVs of individual SSc codes are listed in Table 1 and ranged from 75% to 100%. Among them, the PPV of CREST syndrome codes to identify patients with CREST syndrome was 95% (95% CI, 85–98%).

There were 16 cases with SSc-specific ICD-10 codes who were miscoded (Table 2). The most common reasons for excluding a case after medical record review were for a different connective tissue disease (50% [8 of 16]) or localized scleroderma (18.8% [3 of 16]).

The ICD-10 discharge diagnosis code M34.9 (Systemic sclerosis, unspecified) contributed the most confirmed SSc cases (30% [60 of 200]). The ICD-10 discharge diagnosis code M34.8 (Other forms of systemic sclerosis) had the most false-positive results: 43.8% (7 of 16) of the excluded cases were from code M34.8.

During the same period, 172 SSc patients were included in the Toulouse University Hospital SSc cohort following a hospitalization. The initial hospital stay was encoded M34.0, M34.1, M34.8, or M34.9 for 156 out of these patients. Consequently, the sensitivity of the SSc discharge diagnosis codes was 91% (95% CI, 85–94%). The 16 false-negative cases were all coded L94.0 (localized scleroderma).

Discussion

To the best of our knowledge, this is the first study to assess the accuracy of SSc-specific *ICD-10* SSc discharge diagnosis codes in an electronic administrative database. We found the accuracy of electronic medical records for identifying patients with SSc to be excellent in the French hospital database.

While billing code accuracy studies have been performed for other connective tissue diseases,^{12,13} they have been lacking for SSc. One published¹⁴ study reported that the accuracy of *ICD-9* codes for SSc (710.1 code) was 76%. In this study, the true positive cases were defined by the fulfillment of at least one of the three different classification criteria, including the updated 2013 ACR/EULAR criteria.¹⁰ *ICD-9* only has one code for SSc while *ICD-10* has six, this difference could be an explanation why the accuracy may now be higher.

Our study suggests that most of the false-negative cases were coded “localized scleroderma” this suggests that in order to improve the sensitivity of SSc discharge diagnosis code technicians may need further training to better differentiate SSc from localized scleroderma.

Our findings indicate that, given the strong agreement between the *ICD-10* codes and medical records, SSc can confidently be studied using administrative claims data. Our study shows that PPVs of *ICD-10* codes for SSc are comparable to PPVs of common conditions,^{15–17} including cataract, diabetes mellitus, acute myocardial infarction, or stroke (PPV >90% for all). Our study shows that PPVs of *ICD-10* codes for SSc are more reliable than PPVs of other connective tissue diseases,^{18,19} such as systemic lupus erythematosus and dermatomyositis (PPV <90% for all).

Table 2 Reasons for Excluding Patients Identified by Systemic Sclerosis-Specific International Classification of Diseases, Tenth Revision (*ICD-10*) Codes in the French Hospital Database (2010–2017)^a

| Reason | No. | Associated Systemic Sclerosis <i>ICD-10</i> Codes ^b |
|---|-----|--|
| Localized Scleroderma | 3 | M34.0, M34.8, M34.9 |
| Primitive Sjogren's syndrome | 3 | M34.8, M34.9 ^b |
| Undifferentiated Connective Tissue Diseases | 3 | M34.9 ^b |
| Scleredema Adulorum | 2 | M34.8, M34.9 |
| Raynaud Disease | 1 | M34.8 |
| Systemic Lupus Erythematosus | 1 | M34.1 |
| Dermatomyositis | 1 | M34.8 |
| Multiple Sclerosis | 1 | M34.8 |
| HCV-related cryoglobulinemia vasculitis | 1 | M34.8 |
| Total excluded | 16 | |

Notes: ^aThere were 216 patients with a systemic sclerosis-specific code. ^bCodes contributing more than 1 case.

Abbreviation: HCV, hepatitis C virus.

Our Study Has Several Limitations to Consider

It is limited to a single tertiary center, not necessarily reflecting PMSI coding in the entire France. Even if coding rules are national, there may be differences between establishments. The generalizability of this study may be limited because there may be potential differences in coding in a contained healthcare system like Toulouse University Hospital compared with other systems where insurance approval is needed for medications, procedures, and treatments. Future studies will need to assess the accuracy of *ICD-10* codes for SSc in other settings.

Table 1 Validation of International Classification of Diseases, Tenth Revision (*ICD-10*) Codes Used to Identify Systemic Sclerosis Cases in the Toulouse University Hospital in-Patient Database (2010–2017)^a

| Systemic Sclerosis-Related <i>ICD-10</i> Code | Description | No. ^b | Confirmed ^c | PPV, % |
|---|---|------------------|------------------------|--------|
| M34.0 | Progressive systemic sclerosis | 35 | 33 | 94 |
| M34.1 | CR(E)ST syndrome | 56 | 53 | 95 |
| M34.8 | Other forms of systemic sclerosis | 60 | 53 | 88 |
| -M34.8 + J99.1 | -Systemic sclerosis with lung involvement | 26 | 26 | 100 |
| -M34.8 + G73.7 | -Systemic sclerosis with myopathy | 4 | 3 | 75 |
| M34.9 | Systemic sclerosis, unspecified | 65 | 60 | 92 |
| Total | | 216 | 200 | 93 |

Notes: ^aAll Systemic Sclerosis cases came from codes in this table. ^bIdentified through electronic search. ^cConfirmed through medical record review.

Abbreviations: PPV, positive predictive value; CR(E)ST syndrome, combination of calcinosis, Raynaud phenomenon, (o)esophageal dysfunction, sclerodactyly, telangiectasia.

Conclusion

In summary, this study provides information on the accuracy of *ICD-10* discharge diagnostic codes for SSc in the French hospital database. Consequently, this database is a useful tool for assessing the epidemiology and pharmacoepidemiology of SSc. Given the accuracy of SSc codes, claims data appear to be sufficient for identifying SSc patients for electronic medical record studies. These investigations can be valuable, but the intricacies of each disease and database should be carefully considered when designing studies.

Article Summary: Strengths and Limitations of This Study

- An electronic health record database analysis showed that the positive predictive value of SSc discharge diagnosis codes was 93%.
- The sensitivity of the SSc discharge diagnosis codes was 91%.
- These data suggest that *ICD-10* discharge diagnosis codes can be used to identify cases of systemic sclerosis in the French hospital database.

Systemic sclerosis (SSc) is a systemic orphan connective tissue disease which accounts for significant morbidity-mortality worldwide.¹ Despite the increasing number of population-based reports using health databases for research in connective tissue diseases,² few studies have assessed the accuracy of International Classification of Diseases, Tenth Revision (*ICD-10*) diagnostic codes for SSc.^{3,4} Because claims database are primarily used for billing purposes, it is important to validate the use of diagnostic codes for research and ensure that the coding actually reflects the conditions described in medical records.⁵

In this study, we assessed the accuracy of *ICD-10* discharge diagnosis codes to identify SSc, in the French Hospital Database, compared with medical record review as the gold standard using the Toulouse University Hospital database, a tertiary medical center. In this database, patient information is recorded electronically. These findings will help guide future studies aimed at using electronic medical records to perform research on SSc.

Data Statement

The datasets generated during and/or analysed during the current study are available from the corresponding author on request.

Patient and Public Involvement Statement

It was not possible to involve patients or the public in the design, or conduct, or reporting, or dissemination of our research.

Author contributions

GP designed the study; SDAC, HD, PDM, and GP acquired the data; GP and SDAC carried out the data management; GP performed statistical analyses; GP wrote the paper. All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

There is no funding to report.

Disclosure

All authors declare: no support from any organization for the submitted work; no financial relationships with any organizations that might have an interest in the submitted work in the previous 3 years; no other relationships or activities that could appear to have influenced the submitted work.

References

1. Denton CP, Khanna D. Systemic sclerosis. *Lancet Lond Engl*. 2017;390(10103):1685–1699. doi:10.1016/S0140-6736(17)30933-9
2. Aviña-Zubieta JA, Man A, Yurkovich M, Huang K, Sayre EC, Choi HK. Early cardiovascular disease after the diagnosis of systemic sclerosis. *Am J Med*. 2016;129(3):324–331. doi:10.1016/j.amjmed.2015.10.037
3. Jacobsen S, Halberg P, Ullman S. Mortality and causes of death of 344 Danish patients with systemic sclerosis (scleroderma). *Br J Rheumatol*. 1998;37(7):750–755. doi:10.1093/rheumatology/37.7.750
4. Man A, Zhu Y, Zhang Y, et al. The risk of cardiovascular disease in systemic sclerosis: a population-based cohort study. *Ann Rheum Dis*. 2013;72(7):1188–1193. doi:10.1136/annrheumdis-2012-202007
5. Conte C, Vaysse C, Bosco P, et al. The value of a health insurance database to conduct pharmacoepidemiological studies in oncology. *Therapie*. 2019;74(2):279–288. doi:10.1016/j.therap.2018.09.076
6. Moulis G, Lapeyre-Mestre M, Palmaro A, Pugnet G, Montastruc J-L, Sailler L. French health insurance databases: what interest for medical research? *Rev Med Interne*. 2015;36(6):411–417. doi:10.1016/j.revmed.2014.11.009
7. Bezin J, Duong M, Lassalle R, et al. The national healthcare system claims databases in France, SNIIRAM and EGB: powerful tools for pharmacoepidemiology. *Pharmacoepidemiol Drug Saf*. 2017;26(8):954–962. doi:10.1002/pds.4233

8. WHO. ICD-10 online versions. WHO. Available from: <http://www.who.int/classifications/icd/icdonlineversions/en/>. Accessed April 30, 2020.
9. Professionnaliser le codage du PMSI. Gest hosp. September, 2013. Available from: <https://gestions-hospitalieres.fr/professionnaliser-le-codage-du-pmsi/>. Accessed April 30, 2020.
10. van den Hoogen F, Khanna D, Fransen J, et al. 2013 classification criteria for systemic sclerosis: an American college of rheumatology/European league against rheumatism collaborative initiative. *Ann Rheum Dis*. 2013;72(11):1747–1755. doi:10.1136/annrheumdis-2013-204424
11. Caudrelier L, Moulis G, Lapeyre-Mestre M, Sailler L, Pugnet G. Validation of giant cell arteritis diagnosis code in the French hospital electronic database. *Eur J Intern Med*. 2018. doi:10.1016/j.ejim.2018.10.004
12. Cho S-K, Sung Y-K, Choi C-B, Kwon J-M, Lee E-K, Bae S-C. Development of an algorithm for identifying rheumatoid arthritis in the Korean National Health Insurance claims database. *Rheumatol Int*. 2013;33(12):2985–2992. doi:10.1007/s00296-013-2833-x
13. Hanly JG, Thompson K, Skedgel C. Identification of patients with systemic lupus erythematosus in administrative healthcare databases. *Lupus*. 2014;23(13):1377–1382. doi:10.1177/0961203314543917
14. Valenzuela A, Yaqub A, Fiorentino D, Krishnan E, Chung L. Validation of the ICD-9-CM code for systemic sclerosis using updated ACR/EULAR classification criteria. *Scand J Rheumatol*. 2015;44(3):253–255. doi:10.3109/03009742.2015.1008038
15. Khokhar B, Jette N, Metcalfe A, et al. Systematic review of validated case definitions for diabetes in ICD-9-coded and ICD-10-coded data in adult populations. *BMJ Open*. 2016;6(8):e009952. doi:10.1136/bmjopen-2015-009952
16. Cozzolino F, Montedori A, Abraha I, et al. A diagnostic accuracy study validating cardiovascular ICD-9-CM codes in healthcare administrative databases. The umbria data-value project. *PLoS One*. 2019;14(7):e0218919. doi:10.1371/journal.pone.0218919
17. McCormick N, Bhole V, Lacaille D, Avina-Zubieta JA, Quinn TJ. Validity of diagnostic codes for acute stroke in administrative databases: a systematic review. *PLoS One*. 2015;10(8):e0135834. doi:10.1371/journal.pone.0135834
18. Barnado A, Casey C, Carroll RJ, Wheless L, Denny JC, Crofford LJ. Developing electronic health record algorithms that accurately identify patients with systemic lupus erythematosus. *Arthritis Care Res*. 2017;69(5):687–693. doi:10.1002/acr.22989
19. Kwa MC, Ardalan K, Laumann AE, Nardone B, West DP, Silverberg JI. Validation of international classification of diseases codes for the epidemiologic study of dermatomyositis. *Arthritis Care Res*. 2017;69(5):753–757. doi:10.1002/acr.23010

Clinical Epidemiology

Dovepress

Publish your work in this journal

Clinical Epidemiology is an international, peer-reviewed, open access, online journal focusing on disease and drug epidemiology, identification of risk factors and screening procedures to develop optimal preventive initiatives and programs. Specific topics include: diagnosis, prognosis, treatment, screening, prevention, risk factor modification,

systematic reviews, risk & safety of medical interventions, epidemiology & biostatistical methods, and evaluation of guidelines, translational medicine, health policies & economic evaluations. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use.

Submit your manuscript here: <https://www.dovepress.com/clinical-epidemiology-journal>