

Implementing ICD-11 for Hospital Discharge Coding in Chile: Key Findings for Semantic Interoperability

Sergio Guinez-Molinos

sguinez@utalca.cl

Universidad de Talca

Paulina Vera Montecinos

Servicio de Salud del Maule

Jorge Pacheco Jara

Ministerio de Salud

Daily Piedra Suárez

Ministerio de Salud

Paulina Leighton Quinchao

Ministerio de Salud

Jaime González Díaz

Universidad de Talca

Katri Kontio

Pan American Health Organization (PAHO/WHO)

Research Article

Keywords: International Classification of Diseases, Morbidity, Semantic Interoperability, Health Information Systems, Chile, ICD-11

Posted Date: October 29th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10863797/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Goal:

Assess the diagnostic consistency between physicians and coders in the use of the International Classification of Diseases, 11th Revision (ICD-11).

Materials and Methods

A study was conducted at the San Juan de Dios Hospital in Curico between March and June 2025, incorporating ICD-11 coding into the clinical services of medicine and psychiatry. Eight physicians and two coders were trained in ICD-11, and the national Deisy ICD-11 platform was utilised to validate diagnoses. A total of 457 hospital discharges were analysed to compare diagnostic coding performed by physicians, coders, and a gold standard reference.

Results

For the primary diagnosis, agreement with the coder was 96% (439/457) and with the gold standard was 93% (424/457). For secondary diagnoses, agreement exceeded 90% across all evaluated fields. The Cohen's kappa coefficient between physicians and the coder was 0.963. The main errors were due to omissions in post-coordination and incorrect placement of the external cause code. The gold standard reviewed 100% of the records for which information was available.

Conclusions

This pilot study showed high diagnostic agreement among physicians, coders, and the national reference for coding hospital discharges. However, relevant gaps were identified in post-coordination, in the recording of external causes, and in the standardisation of review criteria. These findings provide initial evidence to guide future strategies for implementing ICD-11 in hospital morbidity.

INTRODUCTION

The International Classification of Diseases, 11th Revision (ICD-11), represents a transformation in diagnostic coding and the interoperability of health information systems (1, 2). The ICD-11, unlike the ICD-10, was developed for digital environments and features a more flexible architecture based on root codes, extension codes, and post-coordination. (3), which allows for a more precise description of clinical complexity and improves the standardised exchange of health information (4–6).

The ICD-11 was approved by the World Health Organization in 2019 and officially came into force in 2022 (7). Its implementation in health systems has been gradual and has required technological adjustments,

training for clinicians and coders, integration with electronic records, and the development of institutional and governance frameworks for its adoption (8, 9).

Several countries have assessed their applicability through pilot studies in contexts of morbidity and mortality. In China, a national study across 59 hospitals reported concordance rates of 82.9% with post-coordinated coding and 92.2% at the root code level, highlighting high acceptance following brief training (10). In contrast, experiences in countries such as Malaysia and Iran have revealed challenges associated with coders' limited familiarity with new concepts, the technical complexity of digital tools, and the need to strengthen institutional capacities to ensure effective implementation (9, 11). Kuwait, for its part, demonstrated successful implementation in clinical settings with ongoing support, although it required relevant technical and educational adjustments (12).

In Latin America, Chile began studying the implementation of ICD-11 for mortality. A recent national study showed an acceptable level of agreement between coders and reference standards for coding causes of death (13). However, the application of the ICD-11 in hospital morbidity records has not yet been systematically documented.

In this context, the Pan American Health Organization, the Ministry of Health of Chile, the Maule Health Service, and the University of Talca developed a pilot study to implement the ICD-11 in hospital morbidity through the electronic clinical record at the San Juan de Dios Hospital in Curico. The aim of this study was to compare diagnostic agreement between physicians and coders using ICD-11 and to identify common errors and implementation needs in a real hospital setting.

MATERIALS AND METHODS

Study Design

A cross-sectional pilot study compared diagnostic agreement between physicians and coders regarding ICD-11 use at San Juan de Dios Hospital in Curicó, a public hospital in the regional healthcare network.

Participants and inclusion criteria

All hospital discharges from internal medicine and psychiatry during the pilot were included if they had at least one diagnosis in the electronic health record. Departments were chosen for convenience and feasibility, based on their willingness to participate, discharge volume, and diagnostic diversity, allowing assessment of ICD-11 for both clinical and mental health diagnoses.

Records were not excluded based on age, sex, or diagnosis, except for discharges without available diagnostic information in the enabled platforms. The total analysed sample consisted of 457 consecutive discharges (Table 1).

Intervention and technological tools

The intervention consisted of integrating the official ICD-11 coding tools into the SISMAULE electronic clinical record (Maule Health Information System). The following tools were used: the official WHO Application Programming Interface (API) (4), the ICD-11 Coding Tool, the ICD-11 Browser(6, 14) and the national platform Deisy CIE-11, developed by the University of Talca in collaboration with the Chilean Ministry of Health through the Department of Statistics and Health Information (DEIS). The latter enabled structured review of diagnoses and collaborative assessment among clinicians, institutional coders, and a national reference (the gold standard).

The platform that enabled the training and subsequent analysis of the records (Deisy CIE-11) was developed in the Laravel 11 framework (15) and uses the MVC (Model-View-Controller) architecture (16). It has a web interface for managing and reviewing encoded records (Fig. 1) and an API for importing records from an electronic clinical record. It uses Blade, JavaScript, jQuery, CSS, and Tailwind CSS in the interface, PHP 8.2 for server-side logic, and PostgreSQL for data persistence. During this pilot, no formal external validation of the software was carried out.

Staff training

Before the pilot began, each service coordinator (medicine and psychiatry) received training. Eight physicians and two coders participated. The training combined two in-person introductory sessions on the standard, the use of the ICD-11 Browser, the Coding Tool, and the official API, along with two practical coding activities using real cases with Deisy CIE-11. The content included the structure of ICD-11, the difference between root code and full code, post-coordination, recording the primary diagnosis, external causes and secondary diagnoses, and validation criteria. The training was supplemented with online resources from DEIS.

Coding and review procedures

Information corresponding to diagnoses coded by the doctor was collected, including the primary diagnosis, external cause, and up to 6 secondary diagnoses, in accordance with the official structure of the hospital discharge form in Chile. For each of these elements, the assessment carried out by the institutional coder and by the gold standard was also included, noting the corrected code, the coding concordance (yes/no), the type of error identified — post-coordination, incorrect code, or inappropriate placement — and its respective justification in free text.

Gold standard

The role of the gold standard was assumed by a national reference from the DEIS, with over 20 years of experience in ICD-10 coding, ICD-11 training, and participation in validation processes. The review was conducted on the Deisy ICD-11 platform (Fig. 1) using uniform WHO-defined criteria to evaluate the root

code, post-coordination, the correct placement of the external cause, and the semantic consistency of the diagnosis.

Data analysis

A descriptive and concordance analysis was conducted using Microsoft Excel for the initial review, cleaning, and validation of the database, and RStudio, version 2025.05.0, for statistical analysis, employing the dplyr, ggplot2, stringr, and tidyr libraries. Firstly, frequency tables were constructed to estimate review coverage by diagnosis type. Subsequently, percentage concordance was calculated between the coding performed by the doctor, the review by the institutional coder, and the gold-standard evaluation. Additionally, confusion matrices were created to compare decisions among evaluators, and a qualitative analysis of the most frequent error tags recorded during the validation process was conducted. Complementarily, a concordance analysis by diagnostic chapter was performed based on the root of the ICD-11 code. For the clinical concordance analysis, the base or root code was used, while for the evaluation of post-coordination and technical errors, the full code was considered. Finally, to estimate the agreement between the physicians and the institutional coder, Cohen's Kappa coefficient was calculated.

As a sensitivity analysis, the PABAK statistic (prevalence-adjusted bias-adjusted kappa) was estimated to assess the robustness of the observed agreement against possible effects of prevalence.

Ethical considerations

This study was conducted in accordance with the ethical principles of the Declaration of Helsinki and applicable Chilean regulations governing the confidentiality and protection of health information.

The study was undertaken for institutional quality-improvement purposes, involved no direct patient intervention, and analysed only anonymised hospital discharge records. Institutional approval was obtained from the Chilean Ministry of Health and the Maule Health Service. Given the retrospective use of anonymised routinely collected data and the absence of direct participant involvement, individual informed consent was not required.

For the analysis of the records, SISMAULE exports them with registration numbers, without identifying patients. This export is carried out through a secure SSL (Secure Sockets Layer) channel and with token authentication (17)

RESULTS

Sample

A total of 457 hospital discharges from the internal medicine (n = 279) and psychiatry (n = 178) services at the San Juan de Dios Hospital in Curicó were analysed.

Screening coverage

The gold standard review coverage was 100% in all fields with recorded information. On the other hand, institutional coders achieved an average coverage of 97%, with a lower degree of revision in the main diagnosis and in the external cause (Table 1).

Table 1
Records by type of diagnosis and review by type of reviewer

Diagnostic	Number of records	Reviewed by coder (%)	Reviewed by gold standard (%)
Main diagnostic	457	442 (97%)	457 (100%)
External cause	267	260 (97%)	267 (100%)
Secondary Diagnosis 1	265	256 (97%)	265 (100%)
Secondary Diagnosis 2	166	162 (98%)	166 (100%)
Secondary Diagnosis 3	99	97 (98%)	99 (100%)
Secondary Diagnosis 4	46	46 (100%)	46 (100%)
Secondary Diagnosis 5	1	1 (100%)	1 (100%)

Diagnostic Concordance

The evaluation of the "correct code" shows high agreement between the doctor's recording and the reviewers' validation. For the primary diagnosis, 96% of records were confirmed correct by the institutional coder, and 93% by the gold standard in a second validation. The agreement on secondary diagnoses exceeded 90% across all fields reviewed, with secondary diagnoses 4 and 5 receiving 100% approval from both evaluators. The external cause showed low agreement (5%–6%), attributable to errors in the field's location within the electronic form rather than to erroneous clinical content (Table 2).

In the comparison between the physician and the coder, the observed agreement was 96,7%, with a Cohen's Kappa coefficient of 0,963 corresponding to a "near-perfect" level of agreement, according to the classification of Landis and Koch (12).

Regarding the sensitivity analysis using PABAK, a value of 0,934 was obtained, indicating very high agreement and confirming the robustness of the main outcome. This calculation was made based on the total number of records available, including those not reviewed by the coder, which constitutes a conservative estimate of the level of agreement (18).

Table 2
Level of agreement in the gold standard review

Diagnostic	Number of records	Coder concordance (%)	Gold standard concordance (%)
Main diagnostic	457	439 (96%)	424 (93%)
External cause	267	13 (5%)	17 (6%)
Secondary Diagnosis 1	265	254 (96%)	250 (94%)
Secondary Diagnosis 2	166	161 (97%)	149 (90%)
Secondary Diagnosis 3	99	97 (98%)	95 (96%)
Secondary Diagnosis 4	46	46 (100%)	44 (96%)
Secondary Diagnosis 5	1	100%	1 (100%)

Agreement between the coder and the gold standard

The agreement between the coder's review and the gold standard for the primary diagnosis was assessed, classifying each case as correct or incorrect.

The confusion matrix shows that, out of a total of 457 primary diagnoses (Table 3):

- In 371 cases, both assessors agreed that the diagnosis was correct.
- In 6 cases, both agreed that it was incorrect.
- In 27 cases, the coder classified the diagnosis as correct, whilst the Gold standard considered it incorrect.
- In 53 cases, the coder classified it as incorrect, but the Gold standard considered it correct.

Overall, this represents 83% agreement between the two assessments, with 17,5% discrepancies.

Table 3
Confusion matrix between the gold standard and the coder

Result		Gold standard		
		Incorrect	Correct	Total
Coder	Incorrect	6	53	59
	Correct	27	371	398
	Total	33	424	457

Error analysis

The gold standard identified errors in 33 main diagnoses, 97% of which were related to the omission of post-coordination. The external cause was the field with the highest number of labelled errors (n = 250), mostly due to incorrect code placement according to recording standards, rather than semantic errors. Five minor errors were detected in the secondary diagnoses, also linked to post-coordination.

To illustrate the main types of errors identified, representative examples of the discrepancies observed during the gold standard review are presented. These examples show that the most common errors are omissions in post-coordination, rather than the incorrect selection of the base code (Table 4).

Table 4

Example of discrepancies identified by the Gold standard during the coding review. These examples reinforce the idea that the discrepancies observed relate primarily to technical aspects – in particular, post-coordination – rather than to the identification of the main clinical diagnosis.

Diagnostic	Medical Code	Code gold standard	Error type	Explanation
Type 2 diabetes mellitus with ketoacidosis	5A11	5A11/5A22.0	Post-coordination	Section on ketoacidosis is missing
Bleeding oesophageal varices	DA26.0Z	DA26.0Z/DB93.1	Post-coordination	Correct but incomplete primary code (may be combined with liver cirrhosis)
Unspecified viral pneumonia with chickenpox	CA40.1Z	CA40.1Z/1E90.Z	Post-coordination	It can be administered after chickenpox

DISCUSSION

This study assessed the feasibility of implementing ICD-11 for hospital morbidity data in a Chilean public healthcare facility by integrating coding into the standard workflow and comparing results among physicians, in-house coders, and a national reference standard. Overall, the results show that the operational adoption of ICD-11 is possible in a real-world healthcare setting, but they also reveal that the proper use of its more advanced features, particularly post-coordination, requires specific training, standardised review criteria and sufficiently tested digital tools.

It is worth noting the agreement between clinicians and the gold standard; whilst the agreement between clinicians and the institutional coder was very high, with a Cohen’s Kappa of 0,63. These results suggest that clinicians can correctly identify the general diagnosis in ICD-11.

The main discrepancies were not related to an incorrect choice of primary diagnosis, but rather to omissions in post-coordination and errors in recording the external cause. This finding is particularly significant because it confirms that the greater complexity

of ICD-11 does not lie solely in recognising a diagnostic category, but in expressing the information with clinical precision through the extensions and combinations permitted by the standard. In other words, the results show that the transition from ICD-10 to ICD-11 cannot be assessed solely by overall agreement but also by the ability to capture greater semantic granularity (19).

Errors associated with post-coordination are, therefore, one of the key lessons learned from the pilot study. In the main assessment, almost all the errors identified by the gold standard fell within this category. From a training perspective, this finding reinforces the need for training programs that go beyond simply navigating the system or searching for terms, and instead delve deeper into the standard's compositional logic, the use of extension codes, and the resolution of real clinical cases with expert feedback.

Another significant finding was the low level of agreement regarding the external cause. However, the qualitative analysis showed that this result was mainly due to errors in the placement of the field within the electronic clinical record, rather than to failures in the clinical identification of the event. This suggests that some of the observed errors depend not only on the user but also on the interface design and the recording workflow.

One of the study's main contributions is to demonstrate that the implementation of ICD-11 should be understood as a socio-technical process, the performance of which depends as much on users' skills as on technological integration, the quality of records and support from technical experts (20).

The results are consistent with the available international literature. As in China and Kuwait, this experience suggests that ICD-11 can be implemented with high levels of concordance when technical, training, and institutional support are in place. At the same time, it is consistent with the findings from Iran and Malaysia in that familiarity with the standard's conceptual framework and the use of specialised digital tools are essential conditions for effective adoption. (21).

Main Limitations

Among the main limitations, it should be noted that the pilot study was conducted in a single hospital and two departments selected on a convenience basis, which limits the generalisability of the findings to other healthcare settings, levels of care, or professional profiles. Similarly, no information was collected on the participants' professional experience, nor were standardised metrics available for usability, time per case or user satisfaction. Nor was formal external validation of the Deisy ICD-11 platform conducted. Consequently, the results should be understood as preliminary evidence of implementation, not as a definitive assessment of the system's performance or of its support tools.

CONCLUSIONS

This pilot study demonstrated a high level of diagnostic concordance in coding hospital discharges using ICD-11, supporting its operational feasibility in the context assessed. However, gaps remain in

post-coordination and the recording of external causes. These findings provide initial evidence to guide future implementation strategies, particularly regarding training, technical support, and the strengthening of coding processes.

Declarations

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest. Some authors were involved in the design and development of the Deisy CIE-11 platform; this involvement did not affect the design, analysis or interpretation of the results.

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

This study was conducted for the purpose of institutional improvement, without direct patient involvement or the collection of personal data, as all records were anonymized. The initiative received institutional approval from the Maule Health Service; informed consent was not required as this was an internal technical validation exercise.

CONSENT FOR PUBLICATION

“Not Applicable”

FUNDING.

Regional Government of Maule: “Transferencia Sistema de Seguimiento de Patologías Crónicas no Transmisibles” (Code: 40.067.967-0).

Acknowledgement

Regional Government of Maule: “Transferencia Sistema de Seguimiento de Patologías Crónicas no Transmisibles” (Code: 40.067.967-0).

Data Availability

The analysed data contain sensitive information and are not publicly available. The R analysis scripts, with no identifiable data, will be made available, upon reasonable request, to the relevant author via an institutional repository following final acceptance of the manuscript.

AUTHORS' CONTRIBUTION

Autor	Contribution
Sergio Guinez-Molinos	Conceptualisation, methodology, initial drafting, manuscript review
Paulina Vera Montecinos	Initial drafting, data analysis, manuscript review
Jorge Pacheco Jara	Coordination, conceptualisation, critical review, final drafting, supervision
Daily Piedra Suárez	Manuscript review, analysis of results
Paulina Leighton	Data analysis, data processing, interpretation of results and drafting of sections of the manuscript.
Jaime González Díaz	Support with data collection, final review
Katri Kontio	Critical review, drafting of the conclusions section

All the authors reviewed and approved the final version

References

1. The Lancet. ICD-11. Lancet. 2019;393(10188):2275. 10.1016/S0140. -6736(19)31205-X PubMed PMID: 31180012.
2. Organización Mundial de la Salud. Clasificación Internacional de Enfermedades Undécima Revisión (CIE-11) [Internet]. 2022 [cited 2024 Dec 23]. Available from: <https://www.who.int/es/news/item/11-02-2022-icd-11-2022-release>
3. Mabon K, Steinum O, Chute CG. Postcoordination of codes in ICD-11. BMC Med Inf Decis Mak. 2021;21. 10.1186/s12911-022-01876-9 . PubMed PMID: 35581649.
4. OMS. API CIE-11 [Internet]. 2025 [cited 2025 Jun 30]. Available from: <https://icd.who.int/icdapi>
5. Harrison JE, Weber S, Jakob R, Chute CG. ICD-11: an international classification of diseases for the twenty-first century. BMC Med Inf Decis Mak BioMed Cent Ltd. 2021. 10.1186/s12911-021-01534-6 . PubMed PMID: 34753471.
6. Chute CG, Çelik C. Overview of ICD-11 architecture and structure. BMC Medical Informatics and Decision Making. BioMed Cent Ltd. 2021. 10.1186/s12911-021-01539-1 . PubMed PMID: 35578335.
7. World Health Organization. ICD-11 Implementation or Transition Guide [Internet]. 2019. Available from:
8. Broadwin C, Zeidan W, Siam M, Kostanjsek N, Doctor HV, Aly EA, et al. Advancing Refugee Health Data Management: The Implementation of ICD-11 in UNRWA's Primary Care System. Int J Environ Res Public Health. 2024;21(9):1121. 10.3390/ijerph21091121.

9. Ooi ECW, Isa ZM, Manaf MRA, Fuad ASA, Ahmad A, Mustapa MN, et al. Factors influencing the intention to use the ICD-11 among medical record officers (MROs) and assistant medical record officers (AMROs) in Ministry of Health, Malaysia. *Sci Rep.* 2024;14(1). 10.1038/s41598-024-60439-2 . PubMed PMID: 38688966.
10. Zhang M, Wang Y, Jakob R, Su S, Bai X, Jing X, et al. Methodologies and key considerations for implementing the International Classification of Diseases-11th revision morbidity coding: insights from a national pilot study in China. *J Am Med Inform Assoc.* 2024;31(5):1084–92. 10.1093/jamia/ocae031 . PubMed PMID: 38427850.
11. Zarei J, Hadianfard AM, Sheikhtaheri A. Preparation for the Implementation of the International Classification of Diseases 11th Revision: A Survey on Clinical Coders' Knowledge of the International Classification of Diseases 11th Revision, Educational Needs, and Readiness to Accept New Roles: A Cross-Sectional Study. *Health Sci Rep.* 2025;8(3).
12. Ibrahim I, Alrashidi M, Al-Salamin M, Kostanjsek N, Jakob R, Azam S, et al. ICD-11 Morbidity Pilot in Kuwait: Methodology and Lessons Learned for Future Implementation. *Int J Environ Res Public Health.* 2022;19(5). 10.3390/ijerph19053057 . PubMed PMID: 35270745.
13. Vera Montecinos P, Pacheco J, Guinez-Molinos S, Piedra D, Velozo L, Fuentes A, et al. Codificación de mortalidad con CIE-11 en Chile: principales resultados y perspectivas para su implementación. *Revista Panam de Salud Pública.* 2025;49:1. 10.26633/RPSP.2025.50.
14. OMS, Docker. ICD API [Internet]. [cited 2025 Jun 30]. Available from: <https://hub.docker.com/r/whoicd/icd-api>
15. Laravel Holdings Inc. Laravel 11.x documentation [Internet]. 2024 [cited 2026 Apr 25]. Available from: <https://laravel.com/docs/11.x>
16. Leff A, Rayfield JT. Web-application development using the model/view/controller design pattern. In: *Proceedings fifth ieee international enterprise distributed object computing conference.* 2001. pp. 118–27.
17. Dierks T, Rescorla E. The transport layer security (TLS) protocol version 1.2 (No. rfc5246) [Internet]. 2008 [cited 2026 Apr 25]. Available from: <https://www.rfc-editor.org/rfc/rfc5246>
18. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol.* 1993;46(5):423–9. 10.1016/0895-4356. (93)90018-v PubMed PMID: 8501467.
19. Fung KW, Xu J, McConnell-Lampsey S, Pickett D, Bodenreider O. A practical strategy to use the ICD-11 for morbidity coding in the United States without a clinical modification. *J Am Med Inform Assoc.* 2023;30(10):1614–21. 10.1093/jamia/ocad. 128 PubMed PMID: 37407272.
20. Boussat B, Jakob R, Boyer L, Romano PS. Strategic pathways to International Classification of Diseases, 11th Revision, adoption in France and the United States. *Health Affairs Scholar.* Oxford University Press; 2025. 10.1093/haschl/qxaf099.
21. Ooi ECW, Isa ZM, Manaf MRA, Fuad ASA, Sidek HF, Ahmad A, et al. Facilitators and challenges to ICD-11 implementation: a qualitative study using the consolidated framework for implementation

Figures

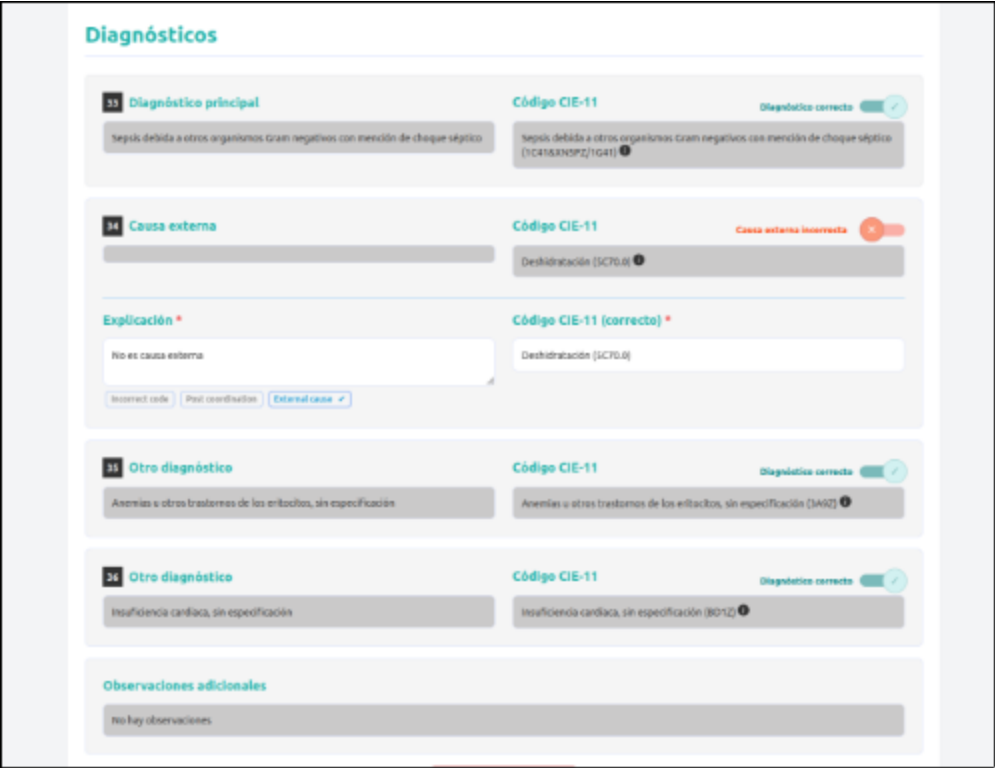


Figure 1

Deisy CIE-11. Record analysis and recoding window for CIE-11. The diagnosis management and validation interface allows the coder to determine whether CIE-11 coding is correct and to record observations for each diagnosis.