**Clinical characteristics and management of patients with secondary hyperparathyroidism undergoing hemodialysis: a feasibility analysis of electronic health records using** **Natural Language Processing**

**SUPPLEMENTAL MATERIALS**

**Supplemental Methods**

# Extracting clinical information from EHRs

The extracted medical terms were organized according to the different EHRs sections (e.g., medical history, laboratory parameters, drug prescriptions, diagnoses, procedures, etc.), and hospital departments. In addition, deep learning classification algorithms based on convolution neural networks (CNN) and word embeddings were used to correct wrongly included occurrences of the clinical terms of interest (false positives) due to ambiguous wording or context. Such algorithms were used to differentiate negative, speculated (possible), and confirmed clinical events such as the combined diagnosis of SHPT and CKD-HD. Additional models were used to properly disambiguate acronyms; for instance, these allowed to differentiate the “hemodialysis” extension of the “HD” acronym from the other 9 possible clinical concepts also referred to as “HD” in Spanish EHRs, including ‘’digestive hemorrhage’’ (hemorragia digestiva, ‘’herniated disk’’ (hernia discal), and “hepatitis D” (hepatitis D). Moreover, for hemodialysis, an extra layer of CNN classification was developed to distinguish “short-term” from “long-term” hemodialysis appearances, thus ensuring the precision of the selected study population by removing false positive cases (see Supplemental Information for details). Medications were identified either as commercial brands or active principles.

# Evaluation of EHRead®’s performance

The goal of this evaluation was to assess the extent to which the (natural language processing) NLP system accurately identified electronic health records (EHRs) that contain mentions of hemodialysis and secondary hyperparathyroidism (SHPT) and consists of a comparison between *EHRead*’s [1-5] reading output and an annotated corpus of EHRs by expert physicians (i.e., ‘gold standard’).

First, this evaluation required the development of an annotated corpus known as *annotation gold standard***.** This is a set of documents marked up with metadata tags related to the study objectives. The development of the gold standard involved the following steps:

* *Text collection*. To determine the size of a corpus that captures the population characteristics as closely as possible while overcoming size limitations, we used Savana’s SampLe Calculator for the Evaluation (SLiCE) tool. This tool calculates the minimum number of annotated EHRs required to obtain the required confidence levels. This calculation is based on the prevalence in the EHRs of the main study variable (in this case, hemodialysis/SHPT). In this case, we aimed for a confidence level of 95% (α = 5%) and interval widths of 10% (percentage points, pp) Thus, SLiCE provided a robust estimation of recall (R) and precision (P) assuring that the true value is at ±5% (pp) with a confidence level of 95%.
* *Annotation task*. To develop the gold standard corpus, a set of documents was first pre-annotated using *EHRead*; these documents only included the main study variables. These documents were then corrected manually using Savana’s Evaluation Tool. The overall goal of this phase is to evaluate the system’s accuracy when identifying records that contain mentions of hemodialysis/SHPT.
* *Annotation of the gold standard*. Two designated medical experts (hereby referred to as ‘the annotators’) at each participating site annotated the set of randomly selected records, always following the annotation guidelines specified by Savana’s medical team. Then, the Inter-Annotator Agreement (IAA) was measured using the F1-Score to ensure the consistency of the guidelines and the reliability of the annotation. The IAA is a metric that indicates the extent to which the different annotators converged in their evaluation, thus also providing information regarding the difficulty of the task. Finally, a third physician acted as judge, reviewing the annotations and resolving any possible discrepancies across annotators. The resulting gold standard corpus served as a resource for the evaluation of EHRead’s performance.
* *Evaluation*. The evaluation of the system is calculated in terms of the standard metrics of recall and precision, and their harmonic mean F1-Score:
* *Recall* = , where *tp* is the number of true positives (i.e., records correctly retrieved), and *fn* is the set of false negatives (i.e., records incorrectly not retrieved). This parameter indicates the amount of information the system retrieves.

* *Precision* = , where *tp* is the number of true positives (i.e., records correctly retrieved) and *fp* is the number of false positives (i.e., records incorrectly retrieved). This parameter indicates the accuracy of the system in retrieving key clinical concepts.

* *F1-Score* = . This parameter gives us an overall performance indicator of information retrieval.

The metrics associated with *EHRead*’s reading performance are shown in **Table S2**.**Supplemental Tables**

Table S1. Hospital sites included in the study

|  |  |  |
| --- | --- | --- |
| **Hospital** | **City** | **Region** |
| Infanta Leonor | Madrid | Madrid |
| Puerta de Hierro | Majadahonda | Madrid |
| La Fe | Valencia | Valencia |
| Son Espases | Palma  | Balearic Islands |
| Infanta Sofía | Madrid | Madrid |
| La Princesa | Madrid | Madrid |
| León | León | Castilla-León |
| Virgen de la Salud | Toledo | Castilla-La Mancha |

Table S2. Performance of EHRead® identifying records that contain key SHPT, hemodialysis, and related variables.

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable** | **Recall** | **Precision** | **F1 score** |
| SHPT | 0.474654 | 0.980952 | 0.639752 |
| Hemodialysis | 0.8125 | 0.984227 | 0.890157 |
| Stroke | 0.745902 | 0.739837 | 0.742857 |
| AMI | 0.6875 | 0.889706 | 0.775641 |
| NSTE ACS | 0.896552 | 0.975 | 0.934132 |
| STE ACD | 0.973684 | 0.936709 | 0.954839 |
| Bone fracture | 0.833333 | 0.675676 | 0.746269 |
| Hypertension | 0.837949 | 0.985525 | 0.905765 |
| Diabetes Mellitus | 0.792308 | 0.956656 | 0.86676 |
| Hypoparathyroidism | 0.822086 | 0.893333 | 0.85623 |
| Parathyroidectomy  | 0.905882 | 0.974684 | 0.939024 |
| Acenocumarol | 0.807692 | 0.830508 | 0.818942 |
| Etelcalcetide | 1 | 0.962963 | 0.981132 |
| Calcitriol | 0.464052 | 0.946667 | 0.622807 |
| Calcium acetate | 0.232143 | 0.464286 | 0.309524 |
| Serum calcium  | 0.658809 | 0.988827 | 0.790767 |
| Serum PTH | 0.642424 | 0.997647 | 0.781567 |
| Serum phosphate | 0.437068 | 0.993711 | 0.607109 |
| SHPT = Secondary hyperparathyroidism; AMI = Acute myocardial infarction; NSTE ACS = Non-ST segment elevation acute coronary syndrome; STE ACS = ST segment elevation acute coronary syndrome; PTH = Parathyroid hormone.  |

**Table S3. Laboratory results at baseline and during follow up.**

|  | BaselineN=282 | 6 monthsN=282 | 12 monthsN=240 |
| --- | --- | --- | --- |
| PTH (pg/ml) |  |  |  |
|  N (%) | 282 (100) | 255 (90.4) | 159 (66.2) |
|  Mean (SD) | 323.4 (213.1) | 274.4 (199.6) | 303.4 (199) |
|  Median (Q1-Q3) | 305 (154.8-436.2) | 226 (128.5-355) | 261 (156.3-381.5) |
|  Missing | 0 | 27 | 81 |
|  PTH controlled (≥150 and ≤450 pg/ml) | 149 (52.8) | 132 (51.8) | 96 (60.4) |
|  PTH > 450 pg/ml | 67 (23.8) | 46 (18) | 29 (18.2) |
|  PTH < 150 pg/ml | 66 (23.4) | 77 (30.2) | 34 (21.4) |
| Ca (mg/dl) |  |  |  |
|  N (%) | 271 (96.1) | 259 (91.8) | 190 (79.2) |
|  Mean (SD) | 8.5 (1.5) | 8.9 (1.0) | 8.7 (1.3) |
|  Median (Q1-Q3) | 8.6 (8.1-9.2) | 9.0 (8.4-9.4) | 8.9 (8.6-9.3) |
|  Missing | 11 | 23 | 50 |
|  Ca controlled (≥8.4 and ≤9.4 mg/dl) | 117 (43.2) | 133 (51.4) | 114 (60) |
|  Ca (> 9.4 mg/dl) | 45 (16.6) | 56 (21.6) | 35 (18.4) |
|  Ca (< 8.4 mg/dl) | 109 (40.2) | 70 (27) | 41 (21.6) |
| P (mg/dl) |  |  |  |
|  N (%) | 160 (56.7) | 136 (48.2) | 100 (41.7) |
|  Mean (SD) | 5.3 (2.4) | 4.7 (1.4) | 4.8 (1.6) |
|  Median (Q1-Q3) | 5.0 (3.9-6.2) | 4.4 (3.6-5.4) | 4.5 (3.8-5.3) |
|  Missing | 122 | 146 | 140 |
|  P controlled (≥2.5 and ≤4.5 mg/dl) | 62 (38.8) | 69 (50.7) | 50 (50.0) |
|  P (> 4.5 mg/dl) | 96 (60) | 62 (45.6) | 47 (47.0) |
|  P (< 2.5 mg/dl) | 2 (1.2) | 5 (3.7) | 3 (3.0) |
| Serum creatinine (mg/dl) |  |  |  |
|  N (%) | 282 (100) | 266 (94.3) | 210 (87.5) |
|  Mean (SD) | 6.3 (2.6) | 6.5 (3.5) | 7.0 (4.3) |
|  Median (Q1-Q3) | 5.5 (4.5-7.1) | 5.9 (4.5-7.6) | 6.3 (4.8-8.3) |
|  Missing | 0 | 16 | 30 |

PTH = Serum Parathyroid hormone; Ca = Serum Calcium; P = Serum Phosphorus.

**Table S4. Medication use at baseline and during follow up. Data expressed as n (%)**

|  | BaselineN=282 | 6 monthsN=282 | 12 monthsN=240 |
| --- | --- | --- | --- |
| SHPT Medication | 209 (74.1) | 187 (66.3) | 143 (59.6) |
| Phosphate Binders | 132(46.8) | 137 (48.6) | 110 (45.8) |
| Sevelamer | 58 (20.6) | 65 (23.0) | 59 (24.6) |
| Lanthanum carbonate | 24 (8.5) | 35 (12.4) | 37 (15.4) |
| Calcium acetate and magnesium carbonate | 13 (4.6) | 13 (4.6) | 10 (4.2) |
| Sucroferric oxyhydroxide | 2 (0.7) | 3 (1.1) | 5 (2.1) |
| Calcium acetate | 20 (7.1) | 22 (7.8) | 17 (7.1) |
| Calcium carbonate | 61 (21.6) | 57 (20.2) | 37 (15.4) |
| Calcimimetics | 27 (9.6) | 31 (11.0) | 29 (12.1) |
| Cinacalcet | 27 (9.6) | 31 (11.0) | 29 (12.1) |
| Etelcalcetide | 1 (0.4) | 1 (0.4) | 1 (0.4) |
| Vitamin D and Analogues | 178 (63.1) | 144 (51.1) | 96 (40.0) |
| Colecalciferol | 58 (20.6) | 53 (18.8) | 24 (10.0) |
| Calcifediol | 81 (28.7) | 62 (22.0) | 47 (19.6) |
| Alfacalcidol | 2 (0.7) | 2 (0.7) | 3 (1.2) |
| Calcitriol | 43 (15.2) | 31 (11) | 20 (8.3) |
| Paricalcitol | 92 (32.6) | 85 (30.1) | 51 (21.2) |
| Anticoagulation | 49 (17.4) | 52 (18.4) | 47 (19.6) |
| Acenocumarol | 49 (17.4) | 52 (18.4) | 47 (19.6) |
| Warfarin | 2 (0.7) | 1 (0.4) | 1 (0.4) |
| Hypertension-related medication | 219 (77.7) | 180 (63.8) | 128 (53.3) |
| Diabetes-related medication | 111 (39.4) | 93 (33.0) | 67 (27.9) |
| Lipid lowering medication | 168 (59.6) | 141 (50) | 105 (43.8) |
|  |

**Supplemental Figures**

**Figure S1. Screened EHRs by hospital department and service.** Total number of EHRs analyzed from 8 hospital sites. Number of records are presented by source, namely admission notes (pink), consultation notes (green), emergency notes (blue), and unknown source (purple), and hospital department/service.

**Figure S2. Structured and unstructured information available in screened EHRs.** Total number of EHRs analyzed from 8 hospital sites (vertical bars). Number of records are presented by the structured nature of the data source, namely laboratory results (“LABORATORY”, structured -laboratory records-; pink), hospital pharmacy records (“PHARMA”, structured -hospital pharmacy; green), and free-text reports (“REPORT”, unstructured; blue).

**REFERENCES**

1. Izquierdo JL, Almonacid C, González Y, Del Rio-Bermúdez C, Ancochea J, Cárdenas R, et al. The Impact of COVID-19 on Patients with Asthma. European Respiratory Journal. 2020:2003142.

2. Hernandez Medrano ITG, J; Belda, C; Urena, A; Salcedo, I; Espinosa-Anke, L; Saggion, H. Savana: Re-using Electronic Health Records with Artificial Intelligence. International Journal of Interactive Multimedia and Artificial Intelligence. 2017;4(7):8-12.

3. Ancochea J, Izquierdo, JL., Medrano, IH., Porras, A., Serrano, M., Lumbreras, S., Del Rio-Bermudez, C., Marchesseau, S., Salcedo, I., Zubizarreta, I., González, Y., Soriano, JB. Evidence of gender differences in the diagnosis and management of COVID-19 patients: an analysis of Electronic Health Records using Natural Language Processing and machine learning. Journal of Women Health. 2020;In press.

4. Izquierdo JL, Ancochea J, Soriano JB. Clinical Characteristics and Prognostic Factors for Intensive Care Unit Admission of Patients With COVID-19: Retrospective Study Using Machine Learning and Natural Language Processing. J Med Internet Res. 2020;22(10):e21801.

5. Graziani D, Soriano JB, Del Rio-Bermudez C, Morena D, Díaz T, Castillo M, et al. Characteristics and Prognosis of COVID-19 in Patients with COPD. Journal of Clinical Medicine. 2020;9(10):3259.