



inDemand: Demand driven co-creation for public entities

CHALLENGE 1: Deep Diver

Pitch

Assistance in the search for diagnoses with suspicion of Professional Illness.

Motivation and description

Work-related illnesses are the leading cause of preventable deaths in developed countries¹. Within them, Professional Diseases (PD) are a group of different pathologies - collected by Spanish law in the Table of Professional Diseases² - whose causal relationship with the job is proven, and which implies a legal responsibility in their prevention, care and benefits. However, most PD go unnoticed as a Common Disease, with the public system assuming the cost of their benefits. Doctors can send suspicions of PD to be assessed by the mutual insurance companies that assume the protection of professional contingencies.

The Murcia Health Service (SMS) has implemented a system of PD suspicion alerts on new diagnoses in the Primary Care (PC) application ([OMI](#)) facilitating the telematic communication of the suspicion to the PD Suspicion Monitoring Unit (USSEP) within the Health Inspection, which processes to the mutual insurance companies suspicions that meet the technical-legal requirements. In the first year of implementation (2015), this regional circuit led to a 68% increase in recognised PD compared to the previous year. However, alerts only act when a new episode is opened in the PC, without acting on historical episodes -where most chronic diseases and cancers are-, or free text within those episodes, or hospital diagnoses.

We can therefore say that this success is only partial as current alerts do not act on most serious diseases. For example, it is estimated that 4% of all cancers could be attributed to occupational exposures. The recognition of professional cancer in Spain is irrelevant, with only a minimal fraction (between 0.1 and 0.2%) of the estimated occupational cancers³.

Main objective

Extend the alerts to detect suspicions of PD by taking advantage of all the information in the clinical history of PC and hospital, as well as the free text of the clinical records, developing an algorithm with a success rate that allows them to be automated.

Pilot functional scope

Approximately 8,000 historical suspicions - of which 4,400 were validated and 3,700 discarded by USSEP - will be used to train and validate the algorithm. All of these patients will include their complete PC and hospital history - PC episodes ([ICPC 2](#) codes), hospital episodes ([ICD 9 MC](#) and [ICD 10](#) codes) - as well as the free text annotated in the PC and hospital records.

After validation of the algorithm, the clinical histories of approximately 200,000 patients from an SMS Health Area will be analyzed, with the same formats as in the previous download used for training and validation of the algorithm.

The downloads will be facilitated by the SMS and no integrations will be required.

Compulsory requirements

1. The solution must be able to identify new suspicions of PD from a list of validated suspicions provided by the SMS.
2. The data sources provided on which this search will be applied will be:
 - a. Episodes of PA coded in ICPC 2.
 - b. Hospital episodes coded in ICD-9 MC and/or ICD-10.

¹ García AM, Gadea R. Incidence and prevalence estimates of work-related diseases in Spain. Aten. Primary 2008;40:439-46.

² As set out in Annex I to Regulation (EC) N^o [R.D. 1299/2006 de 10 de noviembre](#)

³ Incidence data by type of cancer in our country estimated for 2002 by the IARC (IARC-Globocan 2002- <http://www-dep.iarc.fr>)

- c. Free text registered in PC and hospital records, searching for key words (diseases, signs, jobs, exposures...).
 - d. Table of ICD-9 MC and/or ICD-10 diagnostic codes related to PD provided by SMS.
 - e. [INFOCARQUIM](#), information tool on the danger of carcinogens and mutagens and related tumors.
3. The solution will configure a data lake with the raw download of these data sources available to USSEP and SMS. The data will be grouped by patient for their study, seeking to establish algorithms that automate the process and detect new associations of suspected PD.
 4. Applying the validated algorithm to that data lake the company will create a table with the suspicions of PD assigning to each case a probability that will make available to USSEP and SMS.
 5. These suspicions can be filtered by diagnosis and probability of PE to be sent by the USSEP to the family doctors responsible for each patient.
 6. The algorithm will be refined according to the result of the new suspicions generated, adding the evolutionary learning environment to the historical learning environment.

Clinical and Ethical and Data Protection

The Solver undertakes to process the personal data to which it has access as a result of the execution of the contract, observing the principles required by data protection legislation, in particular those relating to data quality, data security and the duty of secrecy, as well as in accordance with the specific instructions received from the data controller, not using the data for any purpose other than the provision of the services described in the object of the contract.

Likewise, it undertakes to observe professional secrecy, maintaining absolute confidentiality and confidentiality on any data that it may come to know on the occasion of the fulfillment of the contract, in accordance with the level of protection established in the [European data protection Regulation \(EU 2016/679\)](#) of the European Parliament and of the Council, of 27 April 2016, relating to the protection of individuals with regard to the processing of personal data and Organic Law 3/2018, of 5 December, on the Protection of Personal Data and guarantee of digital rights, not communicating to any third party the data provided by the data controller. The data controller will determine whether, at the end of the services provided by the data processor, the personal data should be destroyed, returned to the data controller or handed over, where appropriate, to a new data processor.

The destruction of the data will not proceed when there is a legal provision requiring their conservation, in which case they must be returned to the data controller, who will guarantee their conservation for as long as this obligation persists. This obligation will continue even after the end of their relationship with the person in charge.

The Solver shall ensure and be responsible that its employees and/or collaborators receive the data only to the extent that their knowledge is necessary for the provision of the object of the contract.

In the event that the Solver uses the data for purposes other than those stipulated, communicates them or uses them in breach of the instructions set out in this contract, it shall be liable for the infringements set out in Articles 70 et seq. of Organic Law 3/2018, of 5 December, on the Protection of Personal Data and the guarantee of digital rights, in which it has incurred.

Technological

The data lake will be constructed with data from approximately 200.000 patients.

The Solver will build and manage the data lake with historical data extracted from the different data sources. If necessary, data will be anonymized according to mechanisms established by the Challenger. The pilot set up must enable audit capabilities, which at least include user access, task execution and changes in configuration. The anonymized data lake can be hosted by the Solver. If complexity of the connections was too high or privacy could be at risk, the data lake will then be hosted at the Servicio Murciano de Salud. This will be set in a technical session at the beginning of the pilot.

Expected impact and KPI

The result will be the success rate in the alerts offered by the algorithm. Target: area under the ROC curve (AUROC) >0.8

Business opportunity

PD under-registration is universal, being one of the most constant and priority areas of improvement and health savings of any health system, both regional and national. In Spain there is an official PD register ([CEPROSS](#)) in which differences can be found between the different Autonomous Communities, all of which are well below the estimated PD.

The suspicion reporting circuit developed in the Region of Murcia is being replicated in other regions of Spain and the improvement obtained with this challenge would multiply its potency, particularly in the most serious and costly diseases (chronic and cancer). The solution created would allow replicating this improved circuit in any region including other nations, with small adaptations depending on the entities involved.

The savings capacity generated by a PD recognition reaches all health costs (care, pharmacy, interventions, income...). In addition, the right to the benefit generated by PD persists beyond the patient's working life, transcending even his or her heirs in the event of death.