

Harmonization to the OMOIP CDM of cancer patient data at the Modena Oncology Center

Enrico Calanchi¹, Luca Moschetti², Mirko Orsini¹, Laura Delsante¹, Enrica Martinelli³, Andrea Spallanzani², Federica Bertolini², Elisa Pelorelli³, Massimo Dominici³

¹ DataRiver Srl, ² Oncology and Hematology Department, Azienda Ospedaliero-Universitaria di Modena – Policlinico, Modena, Italy, ³ Division of Medical Oncology, Department of Medical and Surgical Sciences for Children and Adults, University Hospital of Modena, Modena, Italy

Background

The evolution of treatments and in particular the development of target therapies and the advent of immunotherapy has revolutionized the therapeutic strategy in the treatment of early and metastatic cancers in the last 10 years. The availability of new treatments has outlined a scenario in which the clinical information relating to the individual patient, is insufficient to define the best treatment but requires constant integration of anatomo-pathological and molecular parameters, obtained through advanced technologies such as gene sequencing. The effectiveness of new drugs, both as monotherapy and in association with chemotherapy or hormone therapy, was correlated with various predictive factors deriving from immunohistochemical or molecular biology analyses. The constant changes in this treatment scenario, determined by registration clinical trials with careful selection of the patients included, determine the need to collect real-practice data to evaluate the feasibility of the aforementioned therapies also in daily outpatient clinical practice. Due to this urgency to provide evidence-based information, the EHDEN (European Health Data & Evidence Network) consortium has launched an extraordinary appeal to invite institutions in Europe to systemically share with all members of the network the data relating to patients suffering from oncological pathologies and treated according to clinical practice. The final aim tends to harmonize the data collected in order to create a common database that allows the rapid extrapolation of high quality evidence. The analysis of the database may allow the identification of any parameters to be able to propose Real World Evidence studies in the future aimed at creating a platform with data, standardized and anonymized according to the computer language used by the platform EHDEN. This evidence will help the scientific community to better understand the impact of treatments on the population, bringing improvements to care in the oncology field also through the conduct of real world studies.

Methods

In this retrospective cohort study, all patients treated with solid tumors starting from the year 2001 at the COM are included. The objective of the study is the creation of a database through the collection of demographic and clinical parameters in order to evaluate the actual outcome of patients treated in normal clinical practice and who do not identify with the subjects enrolled in the clinical studies. This is a single-center, retrospective observational study. For the patients included in the study, clinical data, therapies and procedures performed has been collected retrospectively. In particular, the data of interest are: demographic information (gender, age, place of birth, ethnicity), clinical data (only and exclusively the dates of the start and end of therapies and the dates and results of instrumental re-evaluations will be collected). Histopathological and molecular diagnosis, therapies performed: systemic therapies and locoregional treatments. All patients diagnosed with solid tumors, both prevalent and incident, in the period starting from the year 2001 were enrolled. Eligibility criteria were as follow Inclusion criteria, Age > 18 years, patients diagnosed with solid neoplasia since 2001 under the care of the Modena Oncology Center both in terms of new cases and in terms of patients already under treatment. All disease stages relating to solid tumors that have been the responsibility of oncology for surgery alone, for subsequent therapies or for follow-up. Different databases sources will be used: company, radiological examination, laboratory examination. The systems are integrated and individual patients has been identified by a unique ID code. The IT and Telematics Technologies Service executed patient extraction queries from the various company applications and was responsible for their anonymization. The data has been standardized by the DataRiver using an application provided by the EHDEN consortium, to be made accessible in aggregate form to the members of the network itself. The protocol has been approved by the local ethic committee in January 2024. The analysis, mapping and ETL of the COMNet data has been performed on docker platform hosted on a shared virtual server accessible via VPN to ensure maximum data confidentiality. The environment was based on the Linux operating system Ubuntu 22.04 to hosts the installation of the tools provided by EHDEN. The ETL has been implemented on a PostgreSQL instance to upload the data extracted from the SQL Server database of the information systems in csv format via Python script. Python script has been used also to create a support table for the additional transformations while the most of ETL consists in SQL scripts.

Results

The results highlight the increase in clinical data entry and digitalisation over the last 2 decades. An increase in the use of cancer drugs is observed after 2010, due to the availability of new therapies, especially after the availability of targeted therapies and the development of precision medicine. In addition, a parallel increase in clinical instrumental controls is observed, visible in the increase of measurement and observation information. The lack of data on visits prior to 2010, highlights the change in the method of clinical data collection due to the evolution of the medical record and the progressive digitalisation of the paper data collected in the decades-long history of the oncology center in Modena. (Fig.1 and 2)

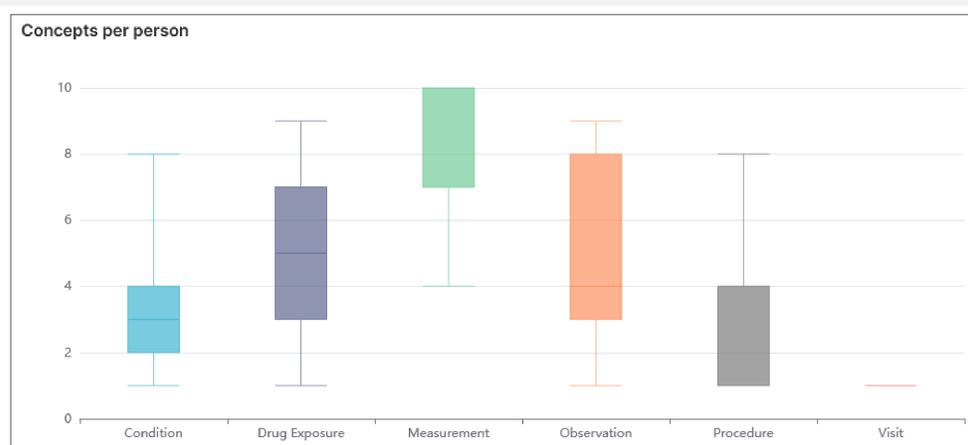


Figure 1

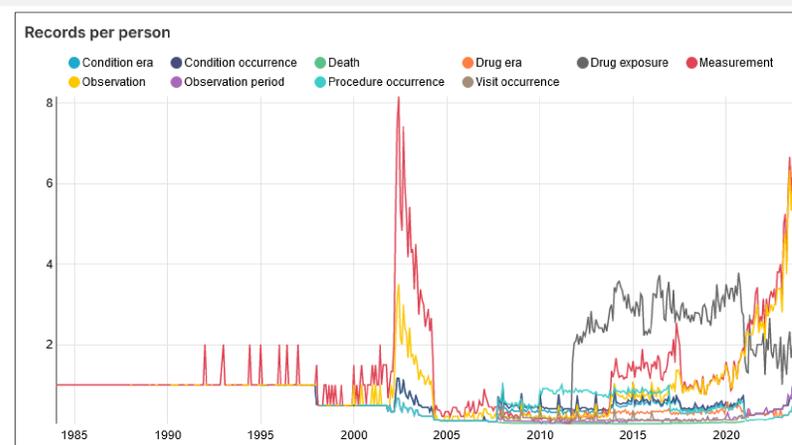


Figure 2

Conclusion

The harmonization process highlighted numerous inconsistencies related to missing dates on information entry or diagnosis of conditions related to patients clinical data stored in the COMNet platform. The analysis of the harmonised data revealed the criticality of data fragmentation due to the incompleteness of the digitalisation process and the presence of different and unconnected software platforms. The objective of achieving greater interoperability and having an integrated view of the information will be one of the guidelines of the re-engineering process of the COMNet platform

