ORACULUM: A retrospective observational epidemiological study using artificial intelligence and natural language processing in electronic health records to characterize the prostate cancer pathway, management and outcomes in Europe, Middle East and Africa (EMEA region)


1Oncology Department, Vall d'Hebron Institute of Oncology, Vall d'Hebron University Hospital, Barcelona, Spain; 2Urology, Hospital Clinic i Provincial de Barcelona, Barcelona, Spain; 3Departments of Surgery and Urology, The Christie and Salford Royal Hospitals, Manchester, UK; 4Radiation Oncology, Hospital Universitario y Politécnico la Fe de Valencia, Valencia, Spain; 5Urology, Uniklinik Köln, Cologne, Germany; 6Urology Department, Hospital de Jerez, Jerez, Spain; 7Medical Oncology Department, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain; 8Dept. Medical Oncology, Hospital Clinico Universitario San Carlos, Madrid, Spain; 9Clinical Research, MedSavana S.L., Madrid, Spain; 10Oncology, Janssen Pharmaceutica, Arles, Arles, France; 11Oncology, Janssen-Cilag Spain, Madrid, Spain; 12Oncology, Janssen-Cilag SA, Madrid, Spain; 13Oncology, Janssen Pharmaceutica NV, Beerse, Belgium; 14Cancer Medicine Department, Institut Gustave Roussy, Villejuif, France

Background: In Prostate Cancer, there is a need for real-world clinical practice data given the high prevalence and incidence of the disease and the rapid changes in treatment options and diagnosis over recent years. Prostate cancer is a scenario in which ‘Big Data’ is particularly applicable because patients have a long disease course, generating an immense amount of data. The outcomes would improve our understanding of the potential implications of different diagnostic/treatment approaches in the different profile of patients suffering from this serious disease in clinical practice. The result of this analysis could also inform the scientific community on new hypotheses for future clinical studies.

Trial design: ORACULUM is a multi-country, multi-centre Artificial Intelligence driven, retrospective, observational study analysing deidentified and aggregated information from original EHRs, in four languages (Spanish, English, French and German). ORACULUM uses SAVANA’s EHRead technology, an innovative data-driven system that applies Artificial Intelligence and Natural Language Program techniques. SAVANA software is able to meaningfully interpret physician notes and numerical values included in clinical records from thousands of patients and translate these into hundreds of variables. ORACULUM uses EHR population-based data since January 2014, to more accurately describe the epidemiology, diagnosis, management, outcomes, prognostic and predictive factors of response of Prostate Cancer (PCa) patients, across all disease stages. ORACULUM may also provide potential correlations that could have remained hidden so far in the existing literature. Here we describe the methodology for data collection and privacy.

Legal entity responsible for the study: Janssen.

Funding: