Data-mining of 110 172 electronic patient records with the ConSoRe tool: An analysis of second primary cancer in a comprehensive cancer center

P. Heudel1, T. Durand1, B. Fervers1, F. Gomez2, M. Rivoire1, T. Bachelor2, L. Claude1, C. Chassagne-Clement1, F. Filieud1, T. Magnetti1, Y. Devaux1, J.-L. Saunierou1, C. Lasset1, D. Perol1, G. Chartiezoff1, C. Pezet1, S. Beaupere1, P. Zourba1, J.-Y. Blay2

1Centre Léon Bérard, Lyon, France; 2Medical oncology, Centre Léon Bérard, Lyon, France

Background: We report a data-mining analysis of 110,172 electronic patient records (EPR) of the Leon Berard Comprehensive Cancer Center (CLB) over a 10 years period to identify characteristics of second primary cancers (SPC).

Methods: ConSoRe is a new generation data analytics solution using natural language processing to search aggregated data and perform advanced data mining. It was used for data extraction from EPR of 110,172 patients (pts), 47,257 men (M) and 62,915 women (W), treated at the CLB from 2007 to 2017. Patient characteristics, treatment and survival were extracted.

Results: Data extraction identified 88,622 pts with at least one cancer. Among them 7,430 (8.4%) had a SPC: 9% (3,475/38,554) in M and 7.9% (3955/50068) in W (p = 3 × 10^-9). Of interest, only 4,296 SPC (37%) were already documented manually by the physicians in the dedicated forms. Mean age at diagnosis of first cancer (FC) is 55.1 years (y) in M and 51.8 y in W. Mean interval (MI) from diagnosis of FC to SPC is 5.3 years (4.4 in M, 6.1 in W; T test p = 4 × 10^-5). Proportions of SPC among specific localization FC are: For M, head & neck cancer (n = 484/5,277, MI: 3.4 y), lymphoma (n = 336/5,611, MI: 5.1 y), prostate cancer (n = 334/4,643, MI: 6.1 y); for W, breast cancer (n = 1,502/21,447, MI: 8.6 y), soft tissue sarcoma (n = 321/3,683, ME: 4.5 y), lymphoma (n = 293/4,432, MI: 6.3 y). Time to SPC differ significantly depending on FC (ANOVA; p = 2 × 10^-14). Pts with SPC were more likely to have received chemotherapy (31.1%) for their FC treatment compared to pts without SPC (20.8%; p = 1.6 × 10^-14). Also, pts with SPC treated by chemotherapy for their FC had shorter MI than pts who did not (3.2 y versus 5.6 y). In multivariate analysis, time from diagnosis of FC to SPC is significantly linked to FC site and shorter when FC was treated with chemotherapy. SPC was significantly correlated to a worth survival: 19% of death with FC at the time of analysis, versus 27.6% of those with SPC (p = 9.7 × 10^-7).

Conclusions: Screening over 100,000 EPR with ConSoRe enabled to retrieve SPC more exhaustively than the physician forms. ConSoRe will be validated in a broader series of 300,000 EPR and used to study SPC risk factors to enable prevention and early detection.

Legal entity responsible for the study: Leon Berard Cancer Center.

Funding: Has not received any funding.

Disclosure: All authors have declared no conflicts of interest.