

ENT COBRA ONTOLOGY: the covariates classification system proposed by the Head & Neck and Skin GEC-ESTRO Working Group for interdisciplinary standardized data collection in head and neck patient cohorts treated with interventional radiotherapy (brachytherapy)

Luca Tagliaferri, MD, PhD¹, Ashwini Budrukhar, MD², Jacopo Lenkowitz, MD³, Mauricio Cambeiro, MD⁴, Francesco Bussu, MD⁵, Jose Luis Guinat, MD⁶, Guido Hildebrandt, MD⁷, Bengt Johansson, MD⁸, Jens E. Meyer, MD⁹, Peter Niehoff, MD¹⁰, Angeles Rovirosa, MD¹¹, Zoltán Takácsi-Nagy, MD¹², Luca Boldrini, MD³, Nicola Dinapoli, MD¹, Vito Lanzotti¹, Andrea Damiani³, Roberto Gatta¹, Bruno Fionda, MD¹, Valentina Lancellotta, MD¹³, Tamer Soror, MD¹⁴, Rafael Martinez Monge, MD⁴, Prof. Vincenzo Valentini, MD³, Prof. György Kovács, MD, PhD¹⁴

¹Fondazione Policlinico Universitario A. Gemelli IRCCS, Roma, Italy, ²Tata Memorial Centre, Mumbai, India, ³Fondazione Policlinico A. Gemelli IRCCS, Università Cattolica del Sacro Cuore, Roma, Italy, ⁴Clinica Universidad de Navarra, Navarra, Spain, ⁵Head of the Otolaryngology Division, Azienda Ospedaliero-Universitaria di Sassari, Italy, ⁶Department of Radiation Oncology, Fundacion Instituto Valenciano de Oncologia, Valencia, Spain, ⁷University Hospital Radiotherapy Department, University of Rostock, Germany, ⁸Department of Oncology, Orebro University Hospital and Orebro University, Sweden, ⁹Head & Neck Surgery Department, AK St. George Hospital, Hamburg, Germany, ¹⁰Department of Radiotherapy, Sana Hospital Offenbach, Offenbach, Germany, ¹¹Hospital Clinic I Universitaria, Barcelona, Spain, ¹²National Cancer Institute, Budapest, Hungary, ¹³Radiation Oncology Section, Department of Surgical and Biomedical Science, University of Perugia and Perugia General Hospital, Perugia, Italy, ¹⁴Interdisciplinary Brachytherapy Unit, University of Lübeck - University Hospital S-H, Campus Lübeck, Germany

Abstract

Purpose: Clinical data collecting is expensive in terms of time and human resources. Data can be collected in different ways; therefore, performing multicentric research based on previously stored data is often difficult. The primary objective of the ENT COBRA (COnsortium for BRachytherapy data Analysis) ontology is to define a specific terminological system to standardized data collection for head and neck (H&N) cancer patients treated with interventional radiotherapy.

Material and methods: ENT-COBRA is a consortium for standardized data collection for H&N patients treated with interventional radiotherapy. It is linked to H&N and Skin GEC-ESTRO Working Group and includes 11 centers from 6 countries. Its ontology was firstly defined by a multicentric working group, then evaluated by the consortium followed by a multi-professional technical commission involving a mathematician, an engineer, a physician with experience in data storage, a programmer, and a software expert.

Results: Two hundred and forty variables were defined on 13 input forms. There are 3 levels, each offering a specific type of analysis: 1. Registry level (epidemiology analysis); 2. Procedures level (standard oncology analysis); 3. Research level (radiomics analysis). The ontology was approved by the consortium and technical commission; an ad-hoc software architecture ("broker") remaps the data present in already existing storage systems of the various centers according to the shared terminology system. The first data sharing was successfully performed using COBRA software and the ENT COBRA Ontology, automatically collecting data directly from 3 different hospital databases (Lübeck, Navarra, and Rome) in November 2017.

Conclusions: The COBRA Ontology is a good response to the multi-dimensional criticalities of data collection, retrieval, and usability. It allows to create a software for large multicentric databases with implementation of specific remapping functions wherever necessary. This approach is well-received by all involved parties, primarily because it does not change a single center's storing technologies, procedures, and habits.

J Contemp Brachytherapy 2018; 10, 3: 260-266

DOI: <https://doi.org/10.5114/jcb.2018.76982>

Key words: data collection, ENT-COBRA, head and neck cancer, GEC-ESTRO, personalized medicine.

Address for correspondence: Jacopo Lenkowitz, MD, Fondazione Policlinico A. Gemelli IRCCS, Università Cattolica del Sacro Cuore, Roma, Italy, Largo Agostino Gemelli 8, 00168 Roma RM, Italy, phone: +34 80095470, e-mail: jacopo.lenkowitz@unicatt.it

Received: 14.02.2018

Accepted: 18.06.2018

Published: 30.06.2018

Purpose

Head and neck squamous cell cancers (HNSCC) currently represent about 4% of the total incidence of malignant diseases in Western countries, with some differences related to the primary site [1,2].

The reduction of behavioral risk factors, such as smoking and alcohol abuse, has indeed validated a decrease in the incidence of head and neck (H&N) malignancies, despite the increase of human papilloma virus (HPV)-related forms [3,4,5,6,7,8].

Some improvements in terms of prognosis have been observed in the last three decades, mainly based on significant technological developments in diagnostic and therapeutic procedures and due to the aforementioned increased incidence of HPV positive forms, which are generally associated with better prognostic outcomes [3,4,8,9,10,11,12,13,14].

Furthermore, the quality of life of HNSCC survivors has significantly improved due to better functional results obtained with the current multimodal oncologic treatments [15,16,17,18,19,20].

The abundance of new options and the progress in individualized treatment procedures have created new challenges for the modern H&N oncology, in which interdisciplinarity has an important role [21,22,23].

In this frame, the possibility to decide a priori which treatment will be the best choice for a specific case, both in terms of modality (surgery versus conservative treatments) and intensity (radiation dose and technique, drug association and surgery extension) is essential.

Limited number of clinical variables that can be managed by physicians in this difficult decision-making process and the absence of dedicated decision support systems (DSS) reflect themselves in the use of generic treatment guidelines. These can hardly take into account other information as staging of disease according to TNM classification of malignant tumors (TNM) and its anatomical primary site [24]. The evidence, on which these guidelines rely on is mainly generated through classical research approaches, preferably by randomized clinical trials (RCT).

To date, randomized clinical trials represent the gold standard for scientific evidence generation, even if their infrastructure still presents some limits, which seem to be tough to manage in the era of personalized medicine. Such kind of study appears cost intensive and usually requires years to be completed. Furthermore, it generally deals only with selected subgroups of patients, hardly meeting all the characteristics presented by general population.

The subgroups of patients selected for RCT often do not represent specific population groups such as elderly, ethnic minorities, socioeconomic categories, or people suffering from specific comorbidities [25,26,27,28].

Additionally, RCT data collection is time consuming procedure, and needs significant efforts in terms of data management and human resources, especially when performed in multi-institutional environments where different data feeding and storage procedures have to be considered.

The combination of the aforementioned vast amount of clinical and molecular data in clinical prediction models,

nomograms [29], and DSS acquires greater importance, and becomes an interesting perspective in the integration of this kind of traditional research [30].

In this frame, the data standardization process could open new possibilities. The overriding position of RCT, gaining importance to reach higher dataset quality, and reduce its entropy – a common ontology table, shared by the different involved institutions could be the solution.

As stated in previous analyses, an ‘ontology’ represents knowledge as a set of concepts within a domain and the relationships between those concepts [30,31]. In practice, an ontology represents a classification system, in which every variable (in this case specifically related to the domain of patients affected by head and neck malignancies) can be represented through uniform and explicit definitions.

The use of ontologies agreed by the different data originating centers can successfully enhance the deep understanding of datasets and the correct utilization of data, as the ontological relationships can address variables defining space (e.g., relationships between institutional and standard terminologies) and time (e.g., consecutive versions of different classification systems).

By this approach, clinical research on H&N cancer will be characterized by a better and less ambiguous understanding of considered variables without differences in storage and interpretation. This data collection model can also increase the number of variables that can be collected over time, comprehending in the dataset all the clinical, therapeutic and technical developments, and updates [32,33]. The quality of collected data can be significantly improved by the standardized data collection (SDC) approach defining which variables should be collected and regulating the most appropriate ways to measure them.

In a previous publication, the characteristics of the COBRA consortium have been described along with the features of the software used to collect and share multi-institutional data [30].

Aim of the present work is to explain and describe the efforts conducted by the Groupe Européen de Curiothérapie-ESTRO Head & Neck and Skin Working Group (GEC-ESTRO H&N WG), in order to define a commonly shared ontology for SDC purposes as well to present the results of the first round of data sharing through the privacy-preserving system, which was developed for this purpose.

Material and methods

The Groupe Européen de Curiothérapie – European Society for Radiotherapy & Oncology Head & Neck and Skin Working Group (GEC-ESTRO H&N and Skin WG) started the H&N COBRA project, approving its structure, and defining its milestones: the consortium agreement, ontology, and the minimal requirements for each center to join the project.

The ontology was defined by a task group (LT, AB, GK), and a technical commission (TeCo) composed by a mathematician (AD), an engineer (RG), a physicist with experience in data storage (JL), a physician with experience in data storage (ND), and a software expert (VL).

This multi-professional group stated the characteristics that the ontology had to meet in order to be accepted.

These requirements included the definition of the data type for each field, the possible values allowed, the cardinality of the items (i.e., single-select or multi-select field), and the allowed range in the case of numerical values.

Also, the taxonomic semantics of particular hierarchical fields were explicit requirements in order to merge into the ontology some semantic annotations that could be extended and exploited for performing inference in the data on a later moment.

From the point of view of semantics, a link to the actual clinical phase was defined for each field in the ontology. For instance, information about a particular treatment (e.g., dose administered, dose rate, date of administration) includes in their meta-data the notion of whether it has been administered for neoadjuvant or adjuvant purposes. In general, all the requirements defined by the consortium aim to define as precise as possible all the knowledge that characterizes the physicians' specialized lexicon (<http://www.openclinical.org/ontologies.html>).

Following the formal definition of the ontology and its requirements, the task group and the TeCo were asked to define the tools for sharing the ontology among the centers in a standardized form. In order to accomplish this task, the "beyond ontology awareness" (BOA) software, capable to reproduce the structure of the ontology, manage the legacy data import, and coordinate data sharing activities was developed [30,34].

All the centers joining the consortium were required to:

1. Install the software with the ontology-compliant case report forms (CRFs) in their servers: the file needed for the installation and setup of both the software and the database, along with a manual to drive the user through this process, were shared among the consortium members via dropbox folders.
2. Upload legacy data into the software's database: a spreadsheet template compliant with the ENT-COBRA ontology definitions was sent to the participants, so that they could fill it with their data. After that, the participants were required to do a simple mapping between this spreadsheet's columns and the actual fields in the ontology database. The mapping and the subsequent data import were performed through BOA.
3. Test the connection to the centralized repository for data sharing: once the legacy data had been imported in the local ontology database, the participants were asked to test their connection to the central repository.
4. Share the data according to the agreed procedure: by means of a dedicated software tool in BOA, the data were anonymized, encrypted, and sent to the central repository via a secure, https-based, web service.

The above-mentioned tasks were performed by designated members of the participant centers with the assistance of the consortium technical team.

Results

Ontology definition and implementation

They were eleven centers of research joining the consortium in total. The characteristics of the ontology were

defined by the TeCo committee and subsequently, they were released an Excel (Excel, Microsoft) template containing all the properties that each field had to be associated with. These properties were defined to be: item name, description label, units of measurement when applicable, item number, response option text when applicable (values of tabular fields), response option value (identifiers of response option text), cardinality of response options (single-select or multi select fields), data type, validation pattern, a flag to state if the field is a required one. Moreover, each field was defined to be associated to the actual therapeutic phase, in which it is recorded: this association is stored in the 'CRF name' property of each field.

Number of the defined variables was 240. Each variable presents 4 properties: name, form, type of field, and levels. Thirteen forms were proposed: 1. Registry and history; 2. Histology; 3. Staging; 4. Protocol; 5. Surgery; 6. External beam radiotherapy; 7. Neoadjuvant chemotherapy (CT); 8. Concomitant CT; 9. Adjuvant CT; 10. Brachytherapy; 11. Follow-up (repeated); 12. Outcome (automatically calculated based on follow-up); 13. Images and treatment files. Field types were: text, number, date, table, files.

The chosen standard file formats are "DICOM" for image and "TXT files" for data treatment. All tables linked with variables are defined. The toxicity is stored with common terminology criteria for adverse events version 4 (CTCAE v. 4.0) scale, and the Radiation Therapy Oncology Group/ EORTC acute and late toxicity scale (RTOG/ EORTC) (for back comparison with retrospective studies). The RTOG/EORTC scale choice was a forced one, as many data are stored using this scale and a direct mapping with CTCAE v. 4.0 is not possible.

There are 3 levels data, each allowing a specific type of analysis: 1. Registry level (epidemiology analysis); 2. Procedures level (standard oncology analysis); 3. Research level (radiomics analysis). The variables of "registry level" are: patient's code, date of birth, gender, ethnicity, age, cancer site, multidisciplinary management, institution, histology type, therapy sequence, death, death date, cause of death. The third level includes image files and all other variables are in the "procedures" level. All the characteristics of the ontology can be found in the supplementary material of this paper.

Software installation and data-sharing

The ontology was proposed by the task force and accepted after internal discussion of the consortium followed by the technical committee (TeCo) approval.

Out of a total number of 11 centers joining the consortium, five had installed the software properly and three of them had successfully imported their legacy data into the COBRA-ontology compliant archive from internal hospital databases (local ethic committee protocol approval was needed).

These centers did not modify their data collection policy, as the software imported the data directly from their database sorting the information according to the ontology. These three centers succeeded in sending anonymized and encrypted data into the cloud-based repository through procedure agreed by the consor-

tium. The first proof run of machine learning analysis took place based on the information coming from these 3 ready centers. Further analyses are scheduled in the next future, aiming to obtain more data from the participating 3 centers and to involve those that will be ready later on.

The three steps of installation, data-import, and data sharing were performed by each center using the guidelines that were provided with the software, and only one remote assistance session was needed by each center per task. No on-site intervention was needed for any of the three tasks and the procedure proved to be straightforward. During the installation procedure in the third center, one problem with the data import was identified. The programmer found a bug in the system and the software code was modified accordingly. No data had been shared before this episode, so that data were correctly imported in the cloud following the consortium rules as well Italian and European laws in respect of the patient's privacy. Whereas local legislation involving patient privacy, data ownership, or other issues make data sharing not feasible, a solution based on PPDM (privacy preserving data mining) [35,36,37] has been developed and implemented. The solution is based on a distributed ecosystem, including preliminary analysis, distributed learning and validation, in a rapid learning framework, enabling researchers to learn and validate predictive models without patient data leaving the institution, where they had been collected in the first place [37].

Epidemiology analysis on the shared data

The first data sharing allowed to run a descriptive analysis of the combined data coming from the three participants centers. To this purpose, a subset of the shared covariates was selected, which included gender, type of histology, date of histology, cancer site, brachytherapy start date, brachytherapy technique. The total number of patients after the data sharing was 325, with center A (Lubeck) contributing 222 patients, center B (Navarra) with 63 patients, and center C (Rome) with 40 patients. The overall dataset contained 10 different histology types, with dates of histology analysis ranging from January 2001 to December 2016. Nineteen different cancer sites were also found and 2 different brachytherapy techniques, with the corresponding brachytherapy start dates ranging from February 2001 to March 2017.

Tables 1-4 show in detail the characteristics of the categorical covariates, namely gender, histology type, cancer site (ID9), and brachytherapy technique, on the overall dataset.

Discussion

These results are certainly encouraging from the perspective of improving clinical research quality, efficiency, and effectiveness. In fact, once the single center is connected through this approach to all the other centers, thanks to the legacy data import procedure, it can continue collecting data in its usual way and at the same time, it is able to participate to larger research projects without sustaining

the related high costs. The data collected in a single center are indeed semi-automatically translated into the ontology and the platform allows to easily select groups of patients and/or groups of covariates to be anonymized, encrypted, and sent to the shared repository, where machine learning algorithms are launched on these large multicentric datasets to build and validate predictive models.

The platform can also allow each center to participate in learning runs in a distributed way, meaning that each center's data stay in the center's own server, and it is the centralized machine learning algorithm to send back and forth the parameters for learning and validating the model until convergence and consensus is reached among participants centers.

Moreover, besides improvements in quality, efficiency, and effectiveness, it is not obvious that a single center has the resources or the will to hire statisticians, physicians, and engineers in order to build software, analyze data, and produce decision support systems (DSS) useful for the clinical practice. Thus, a positive outcome of this effort to produce SDC in a machine learning framework would be gaining the chance to have ready-made, software-embedded, peer-validated predictive models with as low cost as possible for all the centers joining the consortium.

The Authors believe that this ontology is a good answer to a multi-dimensional problem that involves data collection, retrieval, and usability.

Table 1. Number of occurrences and frequencies for 'gender' values in the cloud-shared database

Gender value	# of occurrences	%
Male	226	69.6
Female	98	30.1
Not available	1	0.3

Table 2. Number of occurrences and frequencies for 'histology type' values in the cloud-shared database

Histology type value	# of occurrences	%
Squamous cell carcinoma	272	84
Adenocarcinoma	19	6
Lymphoepithelioma	10	3
Soft tissue tumors and sarcomas	8	2.5
Basal cell neoplasms	7	2.1
Nevi and melanomas	4	1.2
Cystic, mucinous, and serous neoplasms	2	0.6
Blood vessel tumors	1	0.3
Mucoepidermoid neoplasms	1	0.3
Undifferentiated carcinoma	1	0.3
Not available	0	0

Table 3. Number of occurrences and frequencies for 'cancer site (ICD9)' values in the cloud-shared database

Cancer site value (ICD9)	# of occurrences	%
141 Malignant neoplasm of tongue	94	28.9
146 Malignant neoplasm of oropharynx	58	17.8
160 Malignant neoplasm of nasal cavities, middle ear, and accessory sinuses	40	12.3
161 Malignant neoplasm of larynx	35	10.8
147 Malignant neoplasm of nasopharynx	19	5.8
148 Malignant neoplasm of hypopharynx	16	4.9
144 Malignant neoplasm of floor of mouth	14	4.3
145 Malignant neoplasm of other and unspecified parts of mouth	12	3.7
173 Other malignant neoplasm of skin	9	2.8
172 Malignant melanomas of skin	6	1.8
142 Malignant neoplasm of major salivary glands	6	1.8
190 Malignant neoplasm of eye	4	1.2
171 Malignant neoplasm of connective and other soft tissue	3	0.9
143 Malignant neoplasm of gum	2	0.6
199 Malignant neoplasm without specification of site	2	0.6
153 Malignant neoplasm of colon	2	0.6
238 Neoplasm of uncertain behavior of other and unspecified sites and tissues	1	0.3
192 Malignant neoplasm of other and unspecified parts of nervous system	1	0.3
158 Malignant neoplasm of retroperitoneum and peritoneum	1	0.3
Not available	0	0

Multicenter data collection, especially with international project participants, is expensive both in terms of human resources allocation and time spent to perform various tasks. Additionally, national and international law needs to be respected in cases, in which some steps toward process optimization has been taken, and different terms of clinical practice needs to be involved. This is even more accurate, if collected data are used only once and then discarded.

The described procedures with clinical data make them available and reusable also in the future, in different places and by various researchers, and with the

Table 4. Number of occurrences and frequencies for 'brachytherapy technique' values in the cloud-shared database

Brachytherapy technique value	# of occurrences	%
Interstitial	310	95.4
Intracavitary	15	4.6
Not available	0	0

freedom of the use in different research contexts apart of those where they were firstly collected. This also means that more ambitious researches can be conducted, as patients from many institutions are grouped together, the involved numbers can be easily increased by at least one order of magnitude.

The level of the so produced evidence will be defined on a case by case basis, as it will depend on the selected ontologies and on the chosen research setting (e.g., prospective or retrospective).

Furthermore, all the traditional levels of evidence for therapeutic studies could be theoretically met through this approach and modulated by this dynamic framework that will allow the setup of predictive models, whose reliability can be defined by the TRIPOD criteria [38,39].

This new approach will also enable us to conduct researches with more simultaneous parameters, thus answering a pressing issue within the scientific community. The different research contexts mean that our analysis results can be useful for studies that we cannot even imagine today, also giving a practical answer to the great issue of validation, where researchers want to confirm a new model and often struggle to find suitable data.

The potential clinical relevance of this project remains to date not directly foreseeable, as it strongly depends on investigators usage of this new research approach as a practical tool to enhance the quality of the produced scientific evidence enriching the actual power of RCT and allowing to build a more dynamic research context, with clinical data improved by patient reported outcomes obtained through e-medicine techniques and personal devices [40].

Conclusions

The ENT COBRA Ontology represents therefore a good answer to the multi-dimensional issues concerning data collection, retrieval, and usability. This allows to create software for large multicentric databases with the implementation of specific remapping functions. The latter seem to be well received by all involved parties, primarily because this approach does not change the center's storing technologies, procedures, and habits. A further improvement is possible, in which privacy preserving data mining is implemented via distributed learning and validation, enabling learning predictive models without moving patient data from their collecting site.

Disclosure

The authors declare no conflict of interest.

References

- Siegel RL, Miller KD, Jemal A. Cancer statistics, 2017. *CA Cancer J Clin* 2017; 67: 7-30.
- Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer* 2013; 49: 1374-1403.
- Gillison ML, Koch WM, Capone RB et al. Evidence for a causal association between human papillomavirus and a subset of head and neck cancers. *J Natl Cancer Inst* 2000; 92: 709-720.
- Chaturvedi AK, Engels EA, Pfeiffer RM et al. Human papillomavirus and rising oropharyngeal cancer incidence in the United States. *J Clin Oncol* 2011; 29: 4294-4301.
- Sturgis EM, Cinciripini PM. Trends in head and neck cancer incidence in relation to smoking prevalence: an emerging epidemic of human papillomavirus-associated cancers? *Cancer* 2007; 110: 1429-1435.
- Carvalho AL, Nishimoto IN, Califano JA et al. Trends in incidence and prognosis for head and neck cancer in the United States: a site-specific analysis of the SEER database. *Int J Cancer* 2005; 114: 806-816.
- Bussu F, Sali M, Gallus R et al. HPV infection in squamous cell carcinomas arising from different mucosal sites of the head and neck region. Is p16 immunohistochemistry a reliable surrogate marker? *Br J Cancer* 2013; 108: 1157-1162.
- Bussu F, Sali M, Gallus R et al. Human papillomavirus (HPV) infection in squamous cell carcinomas arising from the oropharynx: detection of HPV DNA and p16 immunohistochemistry as diagnostic and prognostic indicators: a pilot study. *Int J Radiat Oncol Biol Phys* 2014; 89: 1115-1120.
- Jemal A, Tiwari RC, Murray T et al. Cancer statistics, 2004. *CA Cancer J Clin* 2004; 54: 8-29.
- Ang KK, Harris J, Wheeler R, et al. Human papillomavirus and survival of patients with oropharyngeal cancer. *N Engl J Med* 2010; 363: 24-35.
- Zur Hausen H, Schulte-Holthausen H, Klein G et al. EBV DNA in biopsies of Burkitt tumours and anaplastic carcinomas of the nasopharynx. *Nature* 1970; 228: 1056-1058.
- Pathmanathan R, Prasad U, Chandrika G et al. Undifferentiated, nonkeratinizing, and squamous cell carcinoma of the nasopharynx. Variants of Epstein-Barr virus-infected neoplasia. *Am J Pathol* 1995; 146: 1355-1367.
- Young LS, Rickinson AB. Epstein-Barr virus: 40 years on. *Nat Rev Cancer* 2004; 4: 757-768.
- Jiong L, Berrino F, Coebergh JW. Variation in survival for adults with nasopharyngeal cancer in Europe, 1978-1989. EURO CARE Working Group. *Eur J Cancer* 1998; 34: 2162-2166.
- Sanguineti G, Geara FB, Garden AS et al. Carcinoma of the nasopharynx treated by radiotherapy alone: determinants of local and regional control. *Int J Radiat Oncol Biol Phys* 1997; 37: 985-996.
- Bussu F, Salgarello M, Adesi LB et al. Oral cavity defect reconstruction using anterolateral thigh free flaps. *B-ENT* 2011; 7: 19-25.
- Bussu F, Paludetti G, Almadori G et al. Comparison of total laryngectomy with surgical (cricohyoidopexy) and nonsurgical organ-preservation modalities in advanced laryngeal squamous cell carcinomas: A multicenter retrospective analysis. *Head Neck* 2013; 35: 554-561.
- Bussu F, Galli J, Valenza V et al. Evaluation of swallowing function after supracricoid laryngectomy as a primary or salvage procedure. *Dysphagia* 2015; 30: 686-694.
- De Vincentiis M, De Virgilio A, Bussu F et al. Oncologic results of the surgical salvage of recurrent laryngeal squamous cell carcinoma in a multicentric retrospective series: emerging role of supracricoid partial laryngectomy. *Head Neck* 2015; 37: 84-91.
- Frakulli R, Galuppi A, Cammelli S et al. Brachytherapy in non-melanoma skin cancer of eyelid: a systematic review. *J Contemp Brachytherapy* 2015; 7: 497-502.
- Bussu F, Tagliaferri L, Mattiucci G et al. Comparison of interstitial brachytherapy and surgery as primary treatments for nasal vestibule carcinomas. *Laryngoscope* 2016; 126: 367-371.
- Tagliaferri L, Bussu F, Rigante M et al. Endoscopy-guided brachytherapy for sinonasal and nasopharyngeal recurrences. *Brachytherapy* 2015; 14: 419-425.
- Tagliaferri L, Bussu F, Fionda B et al. Perioperative HDR brachytherapy for reirradiation in head and neck recurrences: single-institution experience and systematic review. *Tumori* 2017; 103: 516-524.
- Kovács G, Martinez-Monge R, Budrukkar A et al. GEC-ESTRO ACROP recommendations for head & neck brachytherapy in squamous cell carcinomas: 1st update - Improvement by cross sectional imaging-based treatment planning and stepping source technology. *Radiother Oncol* 2017; 122: 248-254.
- Pfister DG, Spencer S, Brizel DM et al. National Comprehensive Cancer Network Clinical Practice Guidelines in Oncology. Head and neck cancers. vers. 1. 2015. Available from URL: http://www.nccn.org/professionals/physician_gls/pdf/head-and-neck.pdf
- Bach PB, Cramer LD, Warren JL et al. Racial differences in the treatment of early-stage lung cancer. *N Engl J Med* 1999; 341: 1198-1205.
- Boyd C, Zhang-Salomons JY, Groome PA et al. Associations between community income and cancer survival in Ontario, Canada, and the United States. *J Clin Oncol* 1999; 17: 2244-2255.
- Hershman D, McBride R, Jacobson JS et al. Racial disparities in treatment and survival among women with early-stage breast cancer. *J Clin Oncol* 2005; 23: 6639-6646.
- Lancellotta V, Kovács G, Tagliaferri L et al. Age Is Not a Limiting Factor in Interventional Radiotherapy (Brachytherapy) for Patients with Localized Cancer. *BioMed Research International* 2018. Article ID 2178469.
- Tagliaferri L, Pagliara MM, Masciocchi C et al. Nomogram for predicting radiation maculopathy in patients treated with Ruthenium-106 plaque brachytherapy for uveal melanoma. *J Contemp Brachytherapy* 2017; 9: 540-547.
- Tagliaferri L, Kovács G, Autorino R et al. ENT COBRA (Consortium for Brachytherapy Data Analysis): interdisciplinary standardized data collection system for head and neck patients treated with interventional radiotherapy (brachytherapy). *J Contemp Brachytherapy* 2016; 8: 336-343.
- Meldolesi E, Balducci M, Chiesa S et al. Perspective of the Large Databases and Ontologic Models of Creation of Pre-clinical and Clinical Results. In: Radiobiology of Glioblastoma: Recent Advances and Related Pathobiology. *Humana Press, Cham* 2016; 293-302.
- Meldolesi E, van Soest J, Alitto AR et al. VATE: VALIDation of high TEchnology based on large database analysis by learning machine. *Colorect Cancer* 2014; 3: 435-450.
- Meldolesi E, van Soest J, Dinapoli N et al. An umbrella protocol for standardized data collection (SDC) in rectal cancer: a prospective uniform naming and procedure convention to support personalized medicine. *Radiother Oncol* 2014; 112: 59-62.
- Tagliaferri L, Gobitti C, Colloca GF et al. A new standardized data collection system for interdisciplinary thyroid cancer management: Thyroid COBRA. *Eur J Intern Med* 2018; 53: 73-78.
- Damiani A. Distributed Learning to Protect Privacy in Multi-centric Clinical Studies. *Artif Intell Med* 2015: 65-75.
- Skripcak T, Belka C, Bosch W et al. Creating a data exchange strategy for radiotherapy research: towards federated databases and anonymized public datasets. *Radiother Oncol* 2014; 113: 303-309.

38. Damiani A. Preliminary data analysis in healthcare multi-centric data mining: a privacy-preserving distributed approach. *J of E-Learning and Knowledge Society* 2018; 14: 71-81.
39. Burns PB, Rohrich RJ, Chung KC. The Levels of Evidence and their role in Evidence-Based Medicine. *Plast Reconstr Surg* 2011; 128: 305-310.
40. Collins GS, Reitsma JB, Altman DG et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *Br J Cancer* 2015; 112: 251-259.