SKIN-COBRA (Consortium for Brachytherapy data Analysis) ontology: The first step towards interdisciplinary standardized data collection for personalized oncology in skin cancer

Valentina Lancellotta, MD, José Luis Guinot, MD, Bruno Fionda, MD, Agata Rembielak, MD, PhD, Alessandro Di Stefani, MD, Stefano Gentileschi, MD, Francesco Federico, MD, Ernesto Rossi, MD, Benjamin Guix, MD, Artur Jan Chyrek, MD, Arenas Meritxell, MD, PhD, Silvia Rodriguez Villalba, MD. Giuseppe Feindinando Collaco, MD, Nicola Dinapoli, MD, PhD, Carlotta Masicocchi, PhD, Jacapo Lenkowicz, PhD, Nicola Dino Capocchiano, PhD, Andrea Damiani, PhD, Vincenzo Valentini, MD, Gyorgy Kovacs, MD, PhD, Luca Tagliaferri, MD, PhD.

1 Fondazione Policlinico Universitario “A. Gemelli” IRCCS, UOC Radioterapia Oncologica, Diagnostica per Immagini, Radioterapia Oncologica ed Ematologia, Rome, Italy. 2 Department of Radiation Oncology, Foundation Instituto Valenciano de Oncologia (I.V.O.), Valencia, Spain. 3 The Christie NHS Foundation Trust, Department of Clinical Oncology, Manchester, United Kingdom. 4 The University of Manchester, Manchester, United Kingdom. 5 Fondazione Policlinico Universitario “A. Gemelli” IRCCS, Istituto di Dermatologia, Roma, Italy. 6 Università Cattolica del Sacro Cuore, Roma, Italy. 7 Fondazione Policlinico Universitario “A. Gemelli” IRCCS, Dipartimento Scienze della Salute della Donna e del Bambino, Unità di Chirurgia Plastica, Roma, Italy. 8 Fondazione Policlinico Universitario “A. Gemelli” IRCCS, Institute of Pathology, Roma, Italy. 9 Fondazione Policlinico Universitario “A. Gemelli” IRCCS, Medical Oncology, Roma, Italy. 10 Foundation IMOR, Department of Radiation Oncology, Barcelona, Spain. 11 Greater Poland Cancer Centre, Brachytherapy Department, Poznań, Poland. 12 Hospital Universitario Sant Joan de Reus, Radiation Oncology Department, Tarragona, Spain. 13 Universitat Rovira i Virgili, Tarragona, Spain. 14 Hospital Clinica Benidorm, Department of Radiation Oncology, Alicante, Spain. 15 University of Lübeck/UKSH-CL, Interdisciplinary Brachytherapy Unit, Lübeck, Germany. 16 Gemelli-INTERACTS, Università Cattolica del Sacro Cuore, Roma, Italia.

Abstract

Purpose: The primary objective of the SKIN-COBRA (Consortium for Brachytherapy data Analysis) ontology is to define a specific terminological system to standardize data collection for non-melanoma skin cancer patients treated with brachytherapy (BT, interventional radiotherapy). Through ontological characterization of information, it is possible to find, isolate, organize, and integrate its meaning.

Material and methods: SKIN-COBRA is a standardized data collection consortium for non-melanoma skin patients treated with BT, including 8 cancer centers. Its ontology was firstly defined by a multicentric and multidisciplinary working group and 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 ninety variables were defined in 10 input forms. There are 3 levels, with 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 technical commission and consortium, and an ad-hoc software system was defined to be implemented in the SKIN-COBRA consortium.

Conclusions: Large databases are natural extension of traditional statistical approaches, a valuable and increasingly necessary tool for modern healthcare system. Future analysis of the collected multinational and multicenter data will show whether the use of the system can produce high-quality evidence to support multidisciplinary management of non-melanoma skin cancer and utilizing this information for personalized treatment decisions.

J Contemp Brachytherapy 2020; 12, 2: 105-110
DOI: https://doi.org/10.5114/jcb.2020.94579

Key words: brachytherapy, ontology, skin cancer.

Address for correspondence: Bruno Fionda, MD, Fondazione Policlinico Universitario “A. Gemelli” IRCCS, UOC Radioterapia Oncologica, Dipartimento di Diagnostica per Immagini, Radioterapia Oncologica ed Ematologia, Largo Agostino Gemelli 8, 00168 Rome, Italy, phone: +39 06 3015 5981, * e-mail: bruno.fionda@yahoo.it

Received: 15.10.2019
Accepted: 17.02.2020
Published: 30.04.2020
Purpose

Non-melanoma skin cancer (NMSC) is the most common type of skin cancer [1,2,3,4,5]. Therapeutic options such as cryotherapy, laser therapy, topical treatment, and photodynamic therapy are usually reserved for an early stage, low-risk, and superficial NMSC, while surgery and radiotherapy remain the mainstays of treatment in high-risk NMSC [6,7,8,9,10,11,12,13,14]. Skin radiotherapy can be delivered by superficial/orthovoltage X-rays, electron beams, megavoltage photons, or brachytherapy (also known as contact radiotherapy and interstitial interventional radiotherapy, BT) [7]. Unfortunately, despite sufficient data in literature, the effects of radiotherapy in NMSC are not evident, as there is only one important publication – an old randomized prospective trial [15,16].

In the past years, external beam radiotherapy (EBRT) and BT procedures became more frequent and are performed with up-to-date technology. Nowadays, multiple imaging modalities are available, which can support target volume definition and may lead to an adequate treatment planning tailoring of volumes in order to spare normal tissues as much as possible [17,18,19,20,21,22]. Differences in optimal treatments between individual patients are not only observed in case of systemic therapies but are also reported in radiotherapy [23,24,25,26]. The concept of “modern personalized medicine” redefines what is currently regarded as “the standard treatment” for each patient. It means that the combination of surgery and radiotherapy could be the right treatment option regarding function and cosmesis preservation. However, new diagnostic tools such as genomic and proteomic profiling as well as analysis of numerical values, and taxonomic semantics. All these items (e.g., single-select or multi-select fields), the range type for each field, the values allowed, the cardinality of choice, can be defined in the SKIN-COBRA ontology.

Differences in optimal treatments between individual patients are not only observed in case of systemic therapies but are also reported in radiotherapy [23,24,25,26]. The concept of “modern personalized medicine” redefines what is currently regarded as “the standard treatment” for each patient. It means that the combination of surgery and radiotherapy could be the right treatment option regarding function and cosmesis preservation. However, new diagnostic tools such as genomic and proteomic profiling as well as analysis of numerical values, and taxonomic semantics. All these items (e.g., single-select or multi-select fields), the range type for each field, the values allowed, the cardinality of choice, can be defined in the SKIN-COBRA ontology.

Material and methods

The structure of the project, the consortium agreement text, and the minimal requirements for each aspirant center were discussed and defined in the frame of the Head & Neck and Skin GEC-ESTRO Working Group (WG). The group also used a GANTT chart to schedule project workload and deadlines [39]. The centers that joined the consortium signed an agreement. The ontology was defined by a specific task group in the frame of the consortium (VL, LT, JLG, AR, GK), and KBO-laboratories (knowledge-based oncology laboratories) technical commission (KBO-TeCo) including a mathematician (AD), a PhD big data engineer (CM), a physicist (JL), a physician expert in big data, a radiomic analysis (ND), and a software expert (NDC). This multidisciplinary team defined characteristics to accept the ontology including the definition of data type for each field, the values allowed, the cardinality of items (e.g., single-select or multi-select fields), the range of numerical values, and taxonomic semantics. All these requirements aim to define the knowledge that characterizes the physicians’ specialized lexicon as precisely as possible (http://www.openclinical.org/ontologies.html).

The ontology was discussed and improved by an independent multidisciplinary external committee (radiation oncologist – BF, BG, AC, MA, SR, AD, VV; dermatologist – ADS; pathologist – FF; medical oncologist – ER; plastic surgeon – SG; geriatric oncologist – GC). As a final step, the ontology was discussed and approved by the consortium. Later, the consortium defined the tools needed to share the ontology among all involved centers in a standardized way. The selected software for this approach was BOA (Beyond Ontology Awareness) (Figure 1). This software is capable of reproducing the structure of ontology, manage the legacy data import wherever needed, and coordinate data sharing activities.

All the centers within the consortium are required to:

1A. Install the service handling the ontology-compliant case report forms (CRFs) on their servers and login into the service through a web browser.

Journal of Contemporary Brachytherapy (2020/volume 12/number 2)
OR

1B. Connect to the centralized repository and manage the CRFs saved on their database directly through a web browser.
2. Upload legacy data into the software’s database.
3. Manually input non-legacy or new data into the desired CRFs.
4. If instruction 1A was chosen, test the connection to the centralized repository to enable data sharing.
5. Share the data according to the agreed procedure.
6. Upload legacy data into the software’s database.
7. Test the connection to the centralized repository for data sharing.
8. Share the data according to the agreed procedure.

Results

Eight centers signed the SKIN-COBRA consortium agreement (Table 1). The Skin Ontology task group, defined by the consortium committee, prepared the above described ontology table proposal (Table 2). After the characteristics of the ontology were defined, an Excel (Microsoft) template was released containing all the properties that each field had to be associated with. These properties included 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. There were 290 of defined variables. Each variable presented 7 properties: description label, left item text, units, response type, response options text, response values or calculations, data type. The variables are organized in 10 forms (Table 3) as follows:

1. Registry and history; 2. Histology; 3. Staging; 4. Surgery; 5. External beam radiotherapy (radical, adjuvant); 6. Systemic treatment (neoadjuvant, concomitant, adjuvant); 7. Brachytherapy (radical, adjuvant); 8. Follow-up; 9. Outcome; 10. Images and treatment files. Field types were defined as: text, number, date, table, files. DICOM

Table 1. SKIN-COBRA consortium requirements

| Requirement                                      |
|-------------------------------------------------|
| Signing of agreement                            |
| Electronic medical record (EMR) for brachytherapy to record patients’ information |
| Possibility to use an ontology-based archives    |
| Possibility to anonymize local data             |
| Possibility to develop advanced multicenter researches |
| Patient’s written informed consent              |

Table 2. SKIN-COBRA framework

| Activity                                                                 |
|--------------------------------------------------------------------------|
| Development and validation of multi-factorial prediction models requiring the availability of large data |
| Each variable must be included into a terminological system              |
| Collected data must be reusable both in time (e.g., in the future) and in the space (across different institutions or research groups) |
| Appropriate mathematical and statistical methods are needed in order to learn from a large collection of data (large database) |
| Patients privacy must be protected, which can be accomplished in two ways, “cloud” and “distributed learning” |
The system guarantees that nothing, except anonymous bounded, lasting through a lifespan of particular study, members; this procedure was only temporary, research the only asset that could be shared among consortium data (Figure 1).

...approaches: local and centralized processing of available ymously processes the clinical data based on two distinct according to the global data dictionary, but if desired, anon-

...facilitates the conversion of legacy pathology archives ac-

...gate the ontology, as the IT architecture of BOA not only (ontology.xls).

...The ontology was available as a supplementary material extra burden for data managers and figures involved.

...archiving, and communication system (PACS), with no extracted from treatment planning system (TPS), picture therapy and imaging information could be automatically - and/or toxicity outcomes or gene expressions. Radio-

...analysis linking to image information with treatment (age, gender, ethnicity, etc.) and used for epidemiological analysis only; 2. The procedure level involved treatment information and oncological outcomes; 3. The research level included the evaluation of diagnostic, treatment, and follow-up imaging available for advanced researches in individualized medicine purposes. Imaging features were developed by several studies such as radiomics analysis linking to image information with treatment and/or toxicity outcomes or gene expressions. Radio-

...therapy and imaging information could be automatically extracted from treatment planning system (TPS), picture archiving, and communication system (PACS), with no extra burden for data managers and figures involved. The ontology was available as a supplementary material (ontology.xls).

**COBRA architecture characteristic**

BOA was chosen in order to implement and propa-
gate the ontology, as the IT architecture of BOA not only facilitates the conversion of legacy pathology archives according to the global data dictionary, but if desired, anony-
mously processes the clinical data based on two distinct approaches: local and centralized processing of available data (Figure 1).

The anonymized cloud-based large database was the only asset that could be shared among consortium members; this procedure was only temporary, research bounded, lasting through a lifespan of particular study. The system guarantees that nothing, except anonymous and non-referable clinical data (with no link to the origi-

...nal local archives) was going to become a part of the large database.

It must also be noted that while generating a predic-
tive model by using information from one or more in-
stitutes, all statistical algorithms would only process the aggregate data, without a need to give other institutes access to individual patient records.

**Discussion**

The results of this project demonstrated that it is possible to create a “common language” (ontology) that connects all participating centers and, through the legacy data import procedure, information can be easily collect-
ed, allowing to take part in larger research projects, even with legacy data. Data collected in every single center is translated into the predefined ontology, which then al-

...ows to easily select groups of patients and/or groups of covariates to be anonymized, encrypted, and sent to the shared repository. The authors believe that this ontology is a viable solution to a multi-dimensional logistic dilemma involving data collection, retrieval, and applicability.

Heterogeneity is a major confounder when interpret-
ing data from skin cancer studies, and two recent NMSC meta-analyses have presented this matter [35,36]. Both studies, despite having the opportunity to evaluate the data from thousands of patients, recommended conduct-
ing prospective trials. SKIN-COBRA has the potential to address this issue.

The challenge, which we aim to overcome through large databases, is to develop multi-factorial predictive models (using historical data to predict future events) to create a decision support system (DSS) to “personalize” the treatment both in terms of efficacy and toxicity [37,38,40,41].

The SKIN-COBRA has some unique features. Our system is absolutely in line with the EU General Data Protection Regulation (GDPR); it may act as a broker, a third-party acting as a connection between the centers participating in a cloud computing service and facilitat-
ing the distribution of data among the centers. It may also be used for implementing distributed learning and can be very helpful for privacy preservation.

The present work proves that the definition of an ontol-
ogy can be used for COBRA as well as for other types of databases. Therefore, the clinician who wants to carry out researchers finds the dataset already configured, with the potential of further customization.

Large databases in radiation oncology mean studying large cohorts of patients and integrating heterogeneous types of data. The science of oncology is moving away from standardized therapies, in which decisions are based only on anatomical, histological, and molecular features that define groups of patients and diseases to personalized approach. This evolution implies a large amount of available data, which will influence physicians in their clinical practice. Furthermore, it allows to choose and combine the best and most up-to-date treatment evidence in a tailored approach. The use of large database is the simplest way to digitally capture a large amount of data regarding patient
characteristics, treatment features, adverse events, and follow-ups. It can be used to generate new knowledge, which may help physicians in this task, providing predictive models and decision support tools (DSS).

By using this DSS it is possible to analyze and interpret the available data to formulate hypotheses and to obtain useful answers for early diagnosis and better treatment. Focusing on the characteristics of individual patient permit to follow the effectiveness, efficiency, and correctness that are typical for current medicine as well as the use of technologies capable of refining and innovating the classic epidemiological methodologies on vast amounts of data [27,28,29]. Promoting and supporting a predictive, preventive, and personalized medicine also means dealing with the costs of healthcare. The data must be of qualitative validity, producing credible results that means dealing with the costs of healthcare. The data must be of qualitative validity, producing credible results that may help physicians in this task, providing predictive models and decision support tools (DSS).

The system could also guide treatment decisions for eventual adverse event management. Additionally, it is useful after the treatment for follow-up and early detection of a relapse. In the future, follow-up needs to integrate all the data collected by wearable devices and connected objects that are being adopted by a large proportion of the population [28]. Continuous and real-time monitoring of abnormal events would lead to earlier detection of relapse and optimization of salvage treatment’s efficiency and cost. Eventually, overall survival will be impacted by such approaches [29].

The combinations of the various parameters can also be represented with nomograms, typical tools that allow the assignment of scores and graphic scales display of the variables that have prognostic relevance.

Radiotherapy is one of the main protagonists of this development, and the number of in-use large databases is increasing fundamentally on the basis of three important phenomena.

The first phenomenon is represented by the digital improvement of diagnostic imaging, which attempts to prevent and early diagnose of the disease; the emergence of new powerful digital diagnostic technologies is visible. The diffusion of these tools has led to the creation of DICOM standard, which defines the guidelines for storing and sharing images. COBRA allows an internal or external validation of the predictive models, so it could provide decision support tools classified also with maximum level of TRIPOD classification [42].

The second phenomenon is connected to the folders and electronic files of patients, since paper files have been eliminated in many centers for a decade or more. The goal is to achieve significant improvements in patient care and not to limit oneself to a mere adoption and compilation of electronic files.

The last phenomenon is characterized by the development of biotechnologies used in the field of the so-called “omics” sciences (genomics, transcriptomics, proteomics, metabolomics, epigenomics, radiomics, etc.). Here, the goal is to provide personalized therapies, which are aimed at the characteristics of individual and inspired by personalized medicine.

Conclusions

Large databases are a natural extension of traditional statistical approaches, a valuable and increasingly necessary tool for a modern healthcare system. The future analysis of the collected multinational and multi-institutional data will show if the use of the system can result in high-quality evidence in the multidisciplinary treatment of NMSC, thus helping to use this information for personalized treatment decisions.

Disclosure

The authors report no conflict of interest.

References

1. Lomas A, Leonardi-Bee J. Bath-Hextall F. A systematic review of worldwide incidence of non-melanoma skin cancer. Br J Dermatol 2012; 166: 1069-1080.
2. Eisemann N, Waldmann A, Geller AC et al. Non-melanoma skin cancer incidence and impact of skin cancer screening on incidence. J Invest Dermatol 2014; 134: 43-50.
3. Katalinic A, Kunze U, Schäfer T. Epidemiology of cutaneous melanoma and non-melanoma skin cancer in Schleswig-Holstein, Germany: incidence, clinical subtypes, tumour stages and localization (epidemiology of skin cancer). Br J Dermatol 2003; 149: 1200-1206.
4. Madan V, Lear JT, Szmolimies RM. Non-melanoma skin cancer. Lancet 2010; 375: 673-685.
5. Apallia Z, Lallas A, Sotiriou E et al. Epidemiological trends in skin cancer. Dermatol Pract Concept 2017; 7: 1-6.
6. Bussu F, Tagliaferri L, Mattiucci G et al. HDR intervention- al radiotherapy (brachytherapy) in the treatment of primary and recurrent head and neck malignancies. Head Neck 2019; 41: 1667-1675.
7. Guinot JL, Rembielak A, Perez-Calatayud J et al. GEC-ESTRO ACROP recommendations in skin brachytherapy. Radiother Oncol 2018; 126: 377-385.
8. Rodriguez S, Arenas M, Gutierrez C et al. Recommendations of the Spanish brachytherapy group (GEB of Spanish Society of Radiation Oncology (SEOR) and the Spanish Society of Medical Physics (SEFM) for high-dose rate (HDR) non melanoma skin cancer brachytherapy. Clin Transl Oncol 2018; 20: 431-442.
9. National Comprehensive Cancer Network Uterine Neoplasms (Version 3 2019): https://www.nccn.org/professionals/physician_gls/pdf/skin.pdf (2019) (accessed: 31 Aug 2019).
10. 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.
11. Lancellotta V, Kovács G, Tagliaferri L et al. The role of personalized Interventional Radiotherapy (brachytherapy) in the management of older patients with non-melanoma skin cancer. J Geriatr Oncol 2018; 10: 514-517.
12. Lancellotta V, Kovács G, Tagliaferri L et al. Age is not a limiting factor in interventional radiotherapy (brachytherapy) for patients with localized cancer. Biomed Res Int 2018; 2018: 2178469.
13. Arenas M, Arguis M, Diez-Presa L et al. Hypofractionated high-dose-rate plesiotherapy in nonmelanoma skin cancer treatment. Brachytherapy 2015; 14: 859-865.
14. Di Stefano A, Del Regno L, Piccerillo A et al. Practical indications for the management of non-melanoma skin cancer patients. G Ital Dermatol Venereol 2017; 152: 264-294.
15. Landthaler M, Braun-Falco O. Anwendung des TDF-Faktors in der Röntgenweichstrahlentherapie. Hautarzt 1989; 40: 774-777.
16. Zaorsky NG, Lee CT, Zhang E, Galloway TJ. Skin cancer brachytherapy vs. external beam radiation therapy (SBRT) meta-analysis. Radiother Oncol 2018; 126: 386-393.
17. Suit H. The Gray Lecture 2001: coming technical advances in radiation oncology. Int J Radiat Oncol Biol Phys 2002; 53: 789-809.
18. Dawson LA, Jaffray DA. Advances in image-guided radiation therapy. J Clin Oncol 2007; 25: 938-946.
19. Thwaites DI, Malicki J. Physics and technology in ESTRO and in radiotherapy and oncology: past, present and into the 4th dimension. Radiother Oncol 2011; 100: 327-332.
20. Rivard MJ, Venselaar JL, Beaulieu L. The evolution of brachytherapy treatment planning. Med Phys 2009; 36: 2136-2153.
21. Haie-Meder C, Siebert FA, Potter R. Image guided, adaptive, accelerated, high dose brachytherapy as model for advanced small volume radiotherapy. Radiother Oncol 2011; 100: 333-343.
22. Arenas M, Sabater S, Sintas A et al. Individualized 3D scanning and printing for non-melanoma skin cancer brachytherapy: a financial study for its integration into clinical workflow. J Contemp Brachytherapy 2017; 9: 270-276.
23. Lambin P, van Stiphout RG, Starmans MH et al. Predicting outcomes in radiation oncology-multifactorial decision support systems. Nat Rev Clin Oncol 2013; 10: 27-40.
24. Sekulic A, Migden MR, Basset-Seguin N et al. Correction to: Accelerated, high dose brachytherapy as model for advanced small volume radiotherapy. Radiother Oncol 2011; 100: 327-332.
25. Migden MR, Rischin D, Schults CD et al. PD-1 blockade for predicting radiation maculopathy in patients treated with Ruthenium-106 plaque brachytherapy for uveal melanoma. J Contemp Brachytherapy 2017; 9: 540-547.
26. Tagliaferri L, Kovács G, Autorino R et al. 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). J Contemp Brachytherapy 2018; 10: 260-266.
27. Tagliaferri L, Budrukker A, Lenkowicz J et al. A new standardized data collection system for interdisciplinary thyroid cancer management: Thyroid COBRA. Eur J Intern Med 2018; 53: 73-78.
28. Wilson JM. Gantt charts: A centenary appreciation. Europ J Oper Res 2003; 149: 430-437.
29. 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.
30. Collins GS, Reitma JB, Altman DG et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ 2015; 350: g7594.