

ENT COBRA (Consortium for Brachytherapy Data Analysis): interdisciplinary standardized data collection system for head and neck patients treated with interventional radiotherapy (brachytherapy)

Luca Tagliaferri, MD, PhD¹, György Kovács, MD, PhD², Rosa Autorino, MD, PhD¹, Ashwini Budrukhar, MD³, Jose Luis Guinot, MD⁴, Guido Hildebrand, MD⁵, Bengt Johansson, MD⁶, Rafael Martinez Monge, MD⁷, Jens E. Meyer, MD⁸, Peter Niehoff, MD⁹, Angeles Rovirosa, MD¹⁰, Zoltán Takácsi-Nagy, MD¹¹, Nicola Dinapoli, MD¹, Vito Lanzotti¹², Andrea Damiani¹³, Tamer Soror², Vincenzo Valentini, MD¹

¹Department of Radiation Oncology - Gemelli-ART, Catholic University, Italy, ²Interdisciplinary Brachytherapy Unit, University of Lübeck - University Hospital S-H, Campus Lübeck, Germany, ³Tata Memorial Centre, Mumbai, India, ⁴Department of Radiation Oncology, Fundacion Instituto Valenciano de Oncologia, Valencia, Spain, ⁵University Hospital Radiotherapy Department, University of Rostock, Germany, ⁶Department of Oncology, Orebro University Hospital and Orebro University, Sweden, ⁷Clinica Universidad de Navarra, Navarra, Spain, ⁸Head & Neck Surgery Department, AK St. George Hospital, Hamburg, Germany, ⁹University Witten Herdecke, Cologne, Germany, ¹⁰Hospital Clinic I Universitari, Barcelona, Spain, ¹¹National Cancer Institute, Budapest, Hungary, ¹²Software programmer manager, KBO-Labs - Gemelli-ART, Catholic University, Italy, ¹³Mathematics; KBO-Labs - Gemelli-ART, Catholic University, Italy

Abstract

Purpose: Aim of the COBRA (Consortium for Brachytherapy Data Analysis) project is to create a multicenter group (consortium) and a web-based system for standardized data collection.

Material and methods: GEC-ESTRO (Groupe Européen de Curiethérapie - European Society for Radiotherapy & Oncology) Head and Neck (H&N) Working Group participated in the project and in the implementation of the consortium agreement, the ontology (data-set) and the necessary COBRA software services as well as the peer reviewing of the general anatomic site-specific COBRA protocol. The ontology was defined by a multicenter task-group.

Results: Eleven centers from 6 countries signed an agreement and the consortium approved the ontology. We identified 3 tiers for the data set: Registry (epidemiology analysis), Procedures (prediction models and DSS), and Research (radiomics). The COBRA-Storage System (C-SS) is not time-consuming as, thanks to the use of "brokers", data can be extracted directly from the single center's storage systems through a connection with "structured query language database" (SQL-DB), Microsoft Access®, FileMaker Pro®, or Microsoft Excel®. The system is also structured to perform automatic archiving directly from the treatment planning system or afterloading machine. The architecture is based on the concept of "on-purpose data projection". The C-SS architecture is privacy protecting because it will never make visible data that could identify an individual patient. This C-SS can also benefit from the so called "distributed learning" approaches, in which data never leave the collecting institution, while learning algorithms and proposed predictive models are commonly shared.

Conclusions: Setting up a consortium is a feasible and practicable tool in the creation of an international and multi-system data sharing system. COBRA C-SS seems to be well accepted by all involved parties, primarily because it does not influence the center's own data storing technologies, procedures, and habits. Furthermore, the method preserves the privacy of all patients.

J Contemp Brachytherapy 2016; 8, 4: 336-343

DOI: 10.5114/jcb.2016.61958

Key words: consortium, data collection, ENT-COBRA, head and neck cancer.

Purpose

Loco-regional recurrence and/or disease progression is the main pattern of failure as well the most common

cause of death in head & neck (H&N) cancer [1,2,3]. The incidence of recurrence after radical treatment may be as high as 30-50% [4,5]. However, H&N cancers can be cured even if a close cooperation among a variety of medical

Address for correspondence: Rosa Autorino, MD, PhD, Department of Radiation Oncology - Gemelli-ART, Catholic University, Pol. A. Gemelli, L.go A. Gemelli 8, 00168 Rome, Italy, phone: +39 320 8652716, fax: +39 06 3015 5339, ✉ e-mail: rosa_autorino@yahoo.it

Received: 19.07.2016

Accepted: 28.07.2016

Published: 24.08.2016

specialists including surgeons, external beam radiotherapy, and interventional radiotherapy (brachytherapy) and medical oncology experts is required to achieve the best outcome [6].

Over the past decade, cancer care has significantly improved, including many new diagnostic methods and treatment modalities [7], which resulted in advances in radiation oncology. Technical developments (especially the involvement of up-to-date imaging methods) have continually improved treatment quality and efficacy also in interventional radiotherapy [8]. On the other hand, the abundance of new options and the progress in individualized medicine has created new challenges. New strategies to improve treatment outcome, including more aggressive therapeutic regimens, have been developed resulting in better results. Unfortunately, the severity and the duration of side effects has also increased at the same time [9].

The choice of treatment of this kind is suggested by general guidelines, which are usually based on evidences of high level clinical research requiring time and finance consumption. Without any doubt, prospective randomized trials (RCTs) play the key role in the definition of clinical guidelines, protocols, and research. However, patients participating in such trials represent a selective subgroup of the general population, resulting in an inherent limiting factor when interpreting results, as the characteristics of the population met in daily clinical practice are very different [10]. Furthermore, some patient groups are under-represented in RCTs, such as the elderly, those with comorbidities [11], or patients with under-represented ethnic and socioeconomic backgrounds [12,13,14]. Therefore, small benefits observed in highly selected trials are likely to disappear when the same treatments are applied in routine practice. Besides RCTs, population-based observational studies are progressively emerging as a complementary form of research, especially to ensure that the results of RCTs translate into tangible benefits when applied to the general population [15]. Observational studies are essential to identify whether clinical practice has changed appropriately, to describe treatment side effects in a wider population, with different age and comorbidities, and to determine whether patients are reaching the desired outcomes with the expected toxicity [16,17,18]. Models for any outcome could benefit from extra information. Therefore, using data of many patients will facilitate building a model also for toxicity [19,20]. However, data collection is time consuming and needs human resources. Often, data are collected with different procedures and it is difficult to perform pooled, multicenter, research based on previously stored multicenter data. Standardized data collection (SDC) improves the quality of the collected data defining variables, which should preferably be collected and regulating how these variables should be measured.

Aim of the COBRA (Consortium for Brachytherapy Data Analysis) project is to create a multicenter group (consortium) and a system for SDC. The long-term aim of this project is the validation of the newest technologies, and the setup of a Decision Support System (DSS) to allow future treatment individualization.

Material and methods

The Groupe Européen de Curiethérapie - European Society for Radiotherapy & Oncology Head and Neck Working Group (GEC-ESTRO H&N WG) started the H&N COBRA project approving its structure and defining: 1) the consortium agreement, 2) the ontology (data-set), 3) minimal requirements for each center to participate in the project.

The WG used the GANTT chart to define work timeline [21]. For every issue, the responsibility and the time to complete the single steps were defined. Every 3 months, a report with the actual status was published on the COBRA web site (<http://www.cobra-brachytherapy.net/Cobra/HOME.html>). The process of standardization of the data collection appears to more effective using a common ontology table. 'Ontology' is a compound word, composed of onto-, from the Greek ὄντος (ontos), which is the present participle of the verb εἶμι (eimi), in other words, 'to be, I am', and λογία (logia), in other words, 'science, study and theory'. Ontology formally represents knowledge as a set of concepts within a domain, and the relationships between those concepts. In practice, an ontology is a classification system where each variable (in this case related to the domain of H&N patients) can be represented by uniform and explicit definitions. Next to changeable definitions, it can define relationships between variables. As these relationships can address variables defining space (e.g. relationships between institutional and standard terminologies) and time (e.g. versions of classifications), ontologies can enhance the understanding of datasets. Eventually, better and unambiguous understanding leads to an approach where H&N cancer research data could be made available without differences in interpretation, today and in the future. This kind of data collection model has to be able also to extend the number of collectable variables over time and to comprehend all the clinical, therapeutic, and technical advances [22].

In the ENT-COBRA project, the ontology was defined by a task group and the consortium evaluated the proposal together with a multi-professional technical commission (TeCo) composed by a mathematician, an engineer, a physician with experience in data storage, a programmer, and a software expert. The minimal requirements for each center to participate in the COBRA consortium are reported in Table 1. The consortium defined the framework of COBRA

Table 1. Minimal requirements of each Centre to participate to the COBRA consortium

In order to participate in the consortium, sign the agreement
To have an Electronic Medical Record (EMR) for brachytherapy to record patient's information
To be able to 'translate' local data into an ontology based archives
To be able to anonymize local data
To use technology able to developed advanced multicentre researches
To provide patient's written informed consents according to local national legislation

software (COBRA framework – see Table 2) and finalized the general ENT COBRA “umbrella” protocol for the approval by local ethics committees. Before sharing the data, each local Ethic Committee had to approve the general umbrella protocol.

Results

The H&N GEC-ESTRO WG approved the project in December 2012 and the text for the agreement was defined in March 2013.

Consortium level

The structure and the rules of the cooperation within the consortium were defined in the agreement text. Each participating center had to indicate a project supervisor in the local unit. The center chief director signed the agreement, designated co-workers authorized to use the COBRA software (maximum 3 per center), and identified a delegate (radiotherapist, surgeon or physicist) for being a part of the ENT-Cobra Executive Committee (ENT-COBRA EC). The ENT-COBRA EC is composed by one representative from each center, and its main aim is to evaluate each application and authorize the participation in the program. The representative is responsible for projects approval and monitoring, for the authorization need of data publication and/or presentation, for the definition of the criteria of author’s name distribution, according to the following principles: the representatives of each participating center have to be responsible for the number of uploaded patient data, for the contribution to data analysis, and for manuscripts editing. The full text

Table 2. COBRA framework

The development and validation of multi-factorial prediction models requires the availability of a large amount of data pathology-bounded considered significant for present and futures studies
Each variable has to be included into a terminological system; adding more variable in the future is possible, if everything about the data is correctly specified (e.g. denomination, measurement units, measurement modality)
Collected data has to be reusable both in time (e.g. in the future) and in the space (across different institutions or research groups); reusability of legacy data is possible, at the condition that suitable semantic remapping functions from old to new data are provided
Appropriate mathematical and statistical methods are needed in order to learn from a large collection of data (Large Database) and help to suggest new modelling hypotheses to be tested
Patients privacy protection has to be protected; this can be accomplished in two ways: <ul style="list-style-type: none"> • by anonymizing data before they leave the collecting institutions walls, making sure that no inverse remapping is available (“cloud” solution) • by exploiting so called “Distributed Learning” solution, in which no data ever leaves the collecting institution but a regressive or classifying predictive model can be learned exactly as if all data had been collected in the same place

of the agreement is available on the COBRA web site. At the time being, eleven centers (10 European and 1 Asian) from 6 countries have signed the agreement.

Ontology level

The ontology was approved by the consortium and by the TeCo, and is composed by 227 variables. Each of these has 4 properties: name, form, type of field, and levels. The variables are arranged in 13 forms (see Table 3). The field types are: text, number, date, table, files. The chosen standard file formats are “DICOM” for image and “TXT files” for data treatment.

Toxicity data have been recorded according to the CTC4 scale as well as with the RTOG scale. RTOG scale was a forced choice because many data had been stored in that form and a direct mapping with CTC4 was not possible. Data are clustered in three tiers:

1. Registry Tier (baseline characteristics): the baseline patient and tumor characteristics that are considered relevant are outlined and organized into the Registry level, the first and most general level that includes the minimal information (age, gender, ethnicity etc.), used for epidemiological analysis only.
2. Procedure Tier (treatment-related characteristics): the baseline treatment and radiotherapy characteristics that are considered relevant have also been defined. These variables are organized in the Procedures Level that includes treatment information with related toxicities, and the evaluation of outcomes in terms of disease free survival and acute and late toxicities. Additional information on radiotherapy will be extracted in an automated way from the record and verified system. More detailed information regarding dosimetric parameters can be calculated using the 3D dose matrix and the imaging information. This information will be retrieved from the PACS system, also in an automated way. This represents no burden to data managers, treating physicians, or patients.

Table 3. Forms

1) Registry and history
2) Histology
3) Staging
4) Protocol
5) Surgery
6) 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

3. Research Tier (imaging) Diagnostic: treatment and follow-up imaging information can be retrieved from the PACS/TPS in an automated way, and organized in the third and most detailed level, the Research level, to be used for advanced research projects. The use and role of medical imaging technologies in clinical oncology have passed from a primarily diagnostic, qualitative, tool to award, a central role in the context of individualized medicine with a quantitative value. Several studies, such as radiomics [23], has been developed to analyze and quantify different imaging features (e.g. descriptors of intensity distribution, spatial relationships between the various intensity levels, texture heterogeneity patterns, descriptors of shape etc.) and the relations of the tumor with the surrounding tissues, to identify a possible relationship between them and treatment outcomes or gene expressions.

COBRA-Storage System level

The COBRA-Storage System (C-SS) architecture was defined having the COBRA framework, the ontology and Ethic Committee (EC) protocols as reference. The software is called BOA (Beyond Ontology Awareness) that is an evolution of SPIDER [24].

Two different strategies will be used depending on the research’s purpose and the centers’ agreement.

Cloud-based large database model

A centralized data record consolidation approach requires a conversion of the data archives according to a global data dictionary. Clinical data are then anonymously re-produced into a cloud-based large database (see Figure 1).

Distributed learning model

A very flexible approach that allows to learn from the data without leaving its center of origin (Figure 2).

The C-SS is not time-consuming, in fact due to the use of “brokers” it can take the data directly from the centers storage systems connecting with SQL, Access®, File Maker Pro® or Excel®. The system is also structured to perform automatic archiving directly from the TPS or after loading machines.

The architecture is based on the concept of “on-purpose data projection”. It means, that a temporary, “virtual”, repository is created “ad hoc” each time, and a new iteration is needed for research purposes. The C-SS architecture is privacy protecting because it will never project data that could identify the individual patient.

Patient’s privacy will also be protected at the architectural level because all data transfer will happen through a fully encrypted pipeline, and data records will be anonymized before leaving the local center’s walls. Mapping between data record and individuals will also be pro-

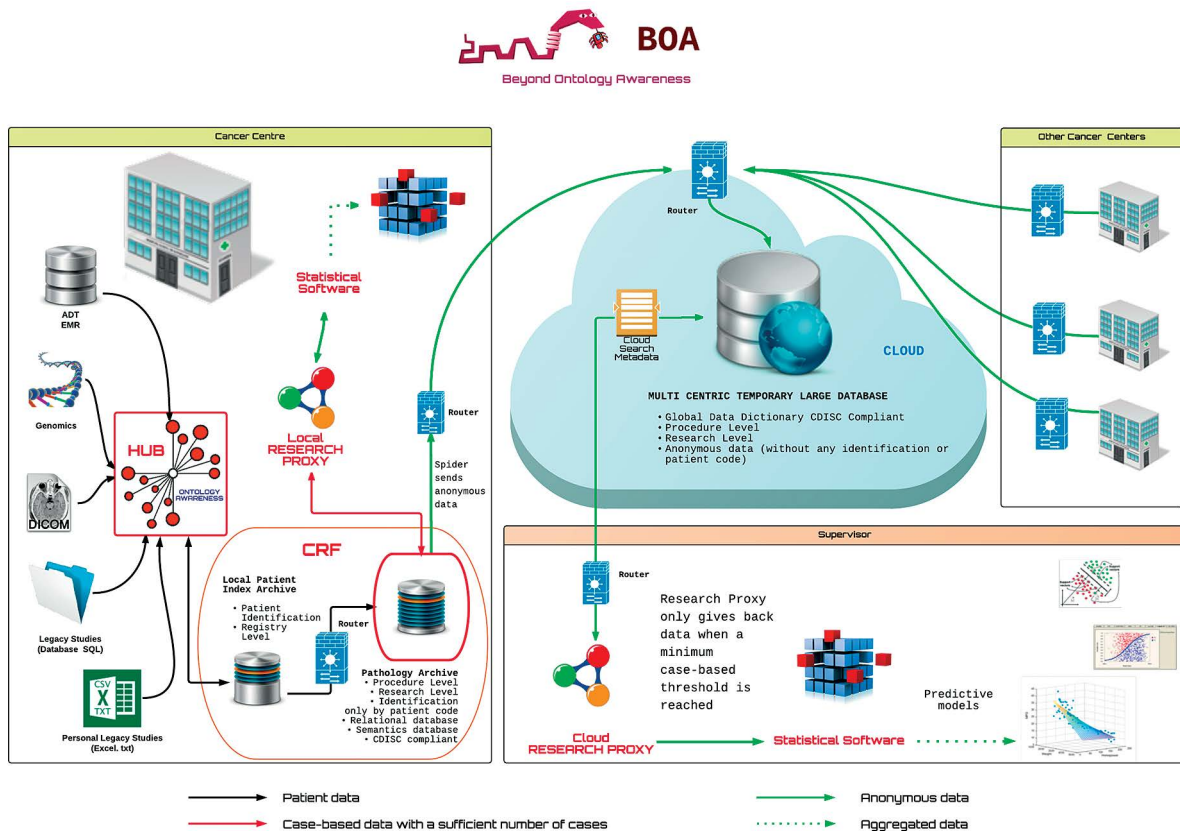


Fig. 1. BOA physically separates privacy relevant information from registry level data splitting this two pieces of information into two databases: “Local Patient Index Archive” and “Pathology Archive”. It sends only clinical data to Cloud Large Database, destroying the inverse mapping, HUB extracts and harmonizes legacy data while making them available for BOA, Local Research Proxy makes local queries on its own pathology database, Cloud Research Proxy run queries on the cloud large database and computes outcomes for each consortium member to use

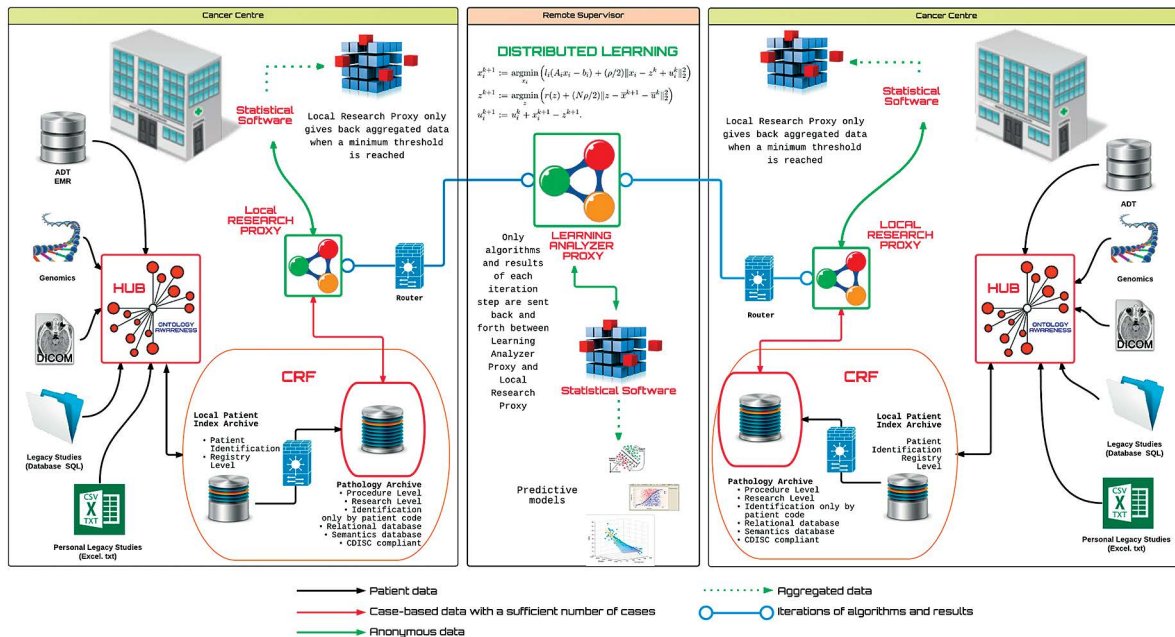


Fig. 2. BOA physically separates privacy relevant information from registry level data splitting this two pieces of information into two databases: "Local Patient Index Archive" and "Pathology Archive". It sends only clinical data to Cloud Large Database, destroying the inverse mapping, HUB (optional module of BOA) extracts and harmonizes legacy data while making them available for BOA, Local Research Proxy (optional module of BOA) makes local queries on its own pathology database. Learning Analyzer Proxy (module of BOA only in distributed mode) sends algorithms directly to Local Research Proxies, taking back from them only the results of each iteration step, with no need to work with shared data in the Cloud anymore. In this mode, Local Research Proxies do not move data around: they only apply iterative algorithms that the Supervisor will use to build consensus and estimate the model's parameters

ected via software procedures, and will never be made available out of the center of origin, thus making virtually useless any attempt of tampering with data transmission and even contacting with the actual data records. This already high degree of protection will be raised even further, where appropriate, through the adoption of secured communication channels (e.g.: virtual private networks over secured connections) and, should necessities arise in order to comply with local regulations or specific policies at the centers' level, decentralized data processing and/or data obfuscation will be added as a further layer of security.

Statistical analysis

Prediction models will be built using two large families of data analysis tools:

1. Inferential regression analysis tools, mainly based on the relationship between outcomes (binary, continuous or multinomial) and covariates, or elements in the dataset, that establish a data-to-outcome one-way link, investigated using traditional statistical tools as linear models, generalized linear models, survival models etc.
2. Machine learning analysis tools, used creating a recursive relationship between outcomes and generating

data, with a complex automation background that can resolve complex relationships between elements in the dataset and final results, too complex in some situations to be investigated using the tools of the first type.

The machine learning approaches can vary but typically are Bayesian networks, Support Vector Machines or Cox regressions. The final model can be presented to the end-user in a variety of ways, such as nomograms, or via interactive websites.

The performance of the models will be assessed in terms of both discrimination and calibration. External validation cohorts will be used for this purpose. Discrimination will be assessed using the c-statistic or area under the curve (AUC) of the receiver operating characteristic (ROC). The c-statistic is comparable to the AUC for dichotomous outcomes but can also be used for Cox regression analyses. Plotting the expected versus the observed outcomes will provide a graphical assessment of the calibration. In addition, the Hosmer-Lemeshow test will be used.

Discussion

The primary and general objective of the COBRA project is to realize a consortium and a system for Standardized Data Collection for Head and Neck cancer patients for the validation of the newest technologies, and to facili-

tate the development of multi-factorial prediction models for different treatment outcomes. The long-term aim is to build a Decision Support System (DSS) based on validated prediction models in order to be able to personalize treatments in terms of both treatment's efficacy and toxicity control. Decision Support System has also the objective to identify patients to be included in future randomized clinical studies, stratifying the different risk classes, depending on the outcomes identified every single time.

Enthusiastic perspectives derived from pre-clinical studies can often influence the adoption of the newest technologies in current brachytherapy practice. On the other hand, the clinical validation of these new technologies can come out difficult because randomized trials comparing different technology levels in treatment approach can be hardly designed, as patients should be assigned to arms with a-priori different technology level. This could result in a conflict with the patients' choices or expectations. Moreover, a long time is usually required for patients recruitment before getting reliable results. The analysis of retrospective case series could be on the other hand a useful tool to obtain data in order to compare different technology levels outcome during a long observation time. It is well known that the comparison of retrospective series can present data collection biases due to the observer known outcome. Those kinds of studies are to be considered always on a lower evidence level when compared to controlled randomized trials. Another problem can derive from the lack of homogeneity in data collection and huge number of parameters that has to be analyzed. The final result is that the clinical evidences of new technologies effectiveness are often inadequate, and strong resistance in novel technology acquisition by multidisciplinary evaluation groups can occur during business management procedures. As new therapeutic strategies and drugs are being tested, it becomes more and more clear that certain subgroups of patients may benefit from a specific treatment, while others will or may even obtain worse outcomes [25]. The same scenario is observed for the toxicity of the treatments, as some patients suffer from severe side-effects while others are relatively unaffected [26]. These observations demonstrate that there is a complex interplay of different factors, which has not yet been deeply investigated. Differences between individual patients are not only observed in the case of different kind of treatments (medication or chemotherapy), but they are also observed in connection with radiotherapy, indicating that the decision to escalate the radiation dose should be individualized. Furthermore, the combination of radiotherapy with surgery could be re-evaluated in order of function- and/or cosmesis preservation. During the last decades, the growth of the power of computer-based analyses has led to access a very large amounts of data in order to find correlations among elements stored in the databases. The possibility to analyze these data can be facilitated through the use of automated procedures that can be guided among pre-defined pathways in order to build up correlations, using Bayesian approaches or support vector machines based analysis software. The amount of available information to explain these observations is enormously expanding due

to new diagnostic tools such as genomics and proteomic profiling (e.g. based on blood or saliva samples), and anatomical and functional imaging techniques (e.g. CT, MRI, PET) that can be used as a starting point to develop predictive models for H&N cancer, useful in offering assistance in clinical decision-making [27,28,29,30].

Response to "Comment on 'Future radiotherapy practice will be based on evidence from retrospective interrogation of linked clinical data sources rather than prospective randomized controlled clinical trials'" [31].

International data-sharing for radiotherapy research: an open-source based infrastructure for multicentric clinical data mining [32]. This knowledge will enable us to predict with greater accuracy the outcome for a specific patient in combination with a certain treatment. It will lead to a clearer identification of risk groups, which could result in stage migration, but it will also stimulate research focused on specific risk groups, trying to find new treatment options or new treatment combinations for these subgroups. It can be expected that in the near future, a treatment will be more personalized, not only preserving patients from unnecessary toxicity and inconveniences, but will enable the choice of the most appropriate treatment. However, a reliable prediction of outcome in order to choose the optimal treatment remains complicated considering the very complex, dynamic, nature of cancer and organs at risk. As an example, a quite recent systematic review concluded that physicians' prediction of survival of terminally ill cancer patients tended to be incorrect in the optimistic direction [33]; similar conclusions were proposed also in another study, which investigated the accuracy of radiation oncologists in survival prediction [34]. Studies investigating the performances of physicians in radiotherapy side effects prediction are currently lacking. However, the ability of human beings (and thus of physicians) to assess the risks and benefits associated with a specific combination of patient, tumor, and treatment characteristics is limited, as it will ultimately include many thousands of parameters.

That is why an appropriate and automated data storage system is encouraged in medical institutions even if data collection needs time and human resources. Unfortunately, data are usually collected differently and it is still very difficult to perform multicenter retrospective researches.

The prospective collection of patient, tumor, and treatment characteristics will facilitate the development of prediction models for survival as well as toxicity outcome, especially through a distributed learning approach and setting up dedicated networks of centers. In addition, data on survival and toxicity can be used to compare results of new emerging radiation delivery techniques, targeted therapies, or chemotherapy regimens after being clinically introduced, with the results obtained of the standard treatment.

The availability of multiple clinical data, together with improved imaging modalities, leads to unprecedented amounts of medical and biological data, which can only be managed using computational methods, not only for static data storage, but also to integrate, analyze, display,

and eventually better understanding. Beside traditional statistical tools (e.g. linear models, generalized linear models, survival models), machine learning appears to be a method for data analysis that automates analytical model building. Using algorithms that iteratively learn from data, machine learning allows computers to find hidden insights without being explicitly programmed where to look. These techniques can overcome problems encountered with conventional statistical methods, especially if data are highly correlated, if there are many variables with a limited number of patients (high-dimensional data), or when many different models have to be tested for their predictive value.

Conclusions

Setting up a consortium showed to be a feasible and practicable tool in the creation of an international and multisystem data sharing system. COBRA C-SS seems to be well accepted by all involved parties, primarily because it does not influence the own data storing technologies, procedures, and habits of the single center. Furthermore, the applied method preserves the privacy of all patient related data at the local user level. The presented multicenter web-based data sharing and the analysis of large amount of data also showed to have a potential role in the validation of the newest diagnostic and therapeutic technologies in the development of multi-factorial prediction models.

Acknowledgement

The authors would like to thank Dr. Luca Boldrini for his support in the revision of the text.

Disclosure

Authors report no conflict of interest.

References

- Vokes EE, Weichselbaum RR. Measurable impact: multimodality therapy of head and neck cancer. *Int J Radiat Oncol Biol Phys* 1993; 27: 481-482.
- Platteaux N, Dirix P, Vanstraelen B et al. Outcome after re-irradiation of head and neck cancer patients. *Strahlenther Onkol* 2011; 187: 23-31.
- Tagliaferri L, Bussu F, Rigante M et al. Endoscopy-guided brachytherapy for sinonasal and nasopharyngeal recurrences. *Brachytherapy* 2015; 14: 419-425.
- Brockstein B, Vokes EE. Concurrent chemoradiotherapy for head and neck cancer. *Semin Oncol* 2004; 31: 786-793.
- Pignon JP, le Maître A, Maillard E et al. Meta-analysis of chemotherapy in head and neck cancer (MACH-NC): an update on 93 randomised trials and 17,346 patients. *Radiother Oncol* 2009; 92: 4-14.
- Kovács G. Modern head and neck brachytherapy: from radium towards intensity modulated interventional brachytherapy. *J Contemp Brachytherapy* 2015; 6: 404-416.
- 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.
- Hoskin PJ, Bownes P. Innovative technologies in radiation therapy: brachytherapy. *Semin Radiat Oncol* 2006; 16: 209-217.
- Bentzen SM, Trotti A. Evaluation of early and late toxicities in chemoradiation trials. *J Clin Oncol* 2007; 25: 4096-4103.
- Zietman AL. Falsification, fabrication, and plagiarism: the unholy trinity of scientific writing. *Int J Radiat Oncol Biol Phys* 2013; 87: 225-227.
- Tyldesley S, Zhang-Salomons J, Groome PA et al. Association between age and the utilization of radiotherapy in Ontario. *Int J Radiat Oncol Biol Phys* 2000; 47: 469-480.
- 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.
- Booth CM, Tannock IF. Randomised controlled trials and population-based observational research: partners in the evolution of medical evidence. *Br J Cancer* 2014; 110: 551-555.
- Pearcey R, Miao Q, Kong W et al. Impact of adoption of chemoradiotherapy on the outcome of cervical cancer in Ontario: results of a population-based cohort study. *J Clin Oncol* 2007; 25: 2383-2388.
- Booth CM. Evaluating patient-centered outcomes in the randomized controlled trial and beyond: informing the future with lessons from the past. *Clin Cancer Res* 2010; 16: 5963-5971.
- Sanoff HK, Carpenter WR, Stürmer T et al. Effect of adjuvant chemotherapy on survival of patients with stage III colon cancer diagnosed after age 75 years. *J Clin Oncol* 2012; 30: 2624-2634.
- Collins GS, Reitsma JB, Altman DG et al. Transparent Reporting of a multivariable prediction model for Individual Prognosis or Diagnosis (TRIPOD). *Ann Intern Med* 2015; 162: 735-736.
- Deasy JO, Bentzen SM, Jackson A et al. Improving normal tissue complication probability models: the need to adopt a "data-pooling" culture. *Int J Radiat Oncol Biol Phys* 2010; 76 (3 Suppl): S151-154.
- Wilson JM. Gantt charts: A centenary appreciation. *Europ J Oper Res* 2003; 149: 430-437.
- 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.
- Parmar C, Leijenaar RT, Grossmann P et al. Radiomic feature clusters and Prognostic Signatures specific for Lung and Head & Neck cancer. *Sci Rep* 2015; 5: 11044.
- Valentini V, Maurizi F, Tagliaferri L et al. Spider: managing clinical data of cancer patients treated through a multidisciplinary approach by a palm based system. *Ital J Public Health JPH* 2008; 6.
- Mok TS, Wu YL, Thongprasert S et al. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med* 2009; 361: 947-957.
- Bentzen SM, Hendry JH. Variability in the radiosensitivity of normal cells and tissues. Report from a workshop organized by the European Society for Therapeutic Radiology and Oncology in Edinburgh, UK, 19 September 1998. *Int J Radiat Oncol Biol Phys* 1999; 75: 513-517.
- 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.
- Kumar V, Gu Y, Basu S, Berglund A et al. Radiomics: the process and the challenges. *Magn Reson Imaging* 2012; 30: 1234-1248.
- Lambin P, Rios-Velazquez E, Leijenaar R et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer* 2012; 48: 441-446.

30. 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.
31. Dekker A. Response to "Comment on 'Future radiotherapy practice will be based on evidence from retrospective interrogation of linked clinical data sources rather than prospective randomized controlled clinical trials'". *Med Phys* 2014; 41: 057102.
32. Roelofs E, Dekker A, Meldolesi E et al. International data-sharing for radiotherapy research: an open-source based infrastructure for multicentric clinical data mining. *Radiother Oncol* 2014; 110: 370-374.
33. Chow E, Harth T, Hruby G et al. How accurate are physicians' clinical predictions of survival and the available prognostic tools in estimating survival times in terminally ill cancer patients? A systematic review. *Clin Oncol (R Coll Radiol)* 2001; 13: 209-218.
34. Chow E, Davis L, Panzarella T et al. Accuracy of survival prediction by palliative radiation oncologists. *Int J Radiat Oncol Biol Phys* 2005; 61: 870-873.