BD2Decide

Big Data and models for personalized Head and Neck Cancer Decision support

TITLE	**User needs and use cases**			
Deliverable No.	D2.1			
EDITOR	AOP - E. Martinelli			
Contributors	All Consortium partners			
WorkPackage No.	WP2	**WorkPackage Title**	Requirements	
Status[1]	Final	**Version No.**	3.0	
Delivery date	01/05/2016	**Actual delivery**	09/05/2016	
Dissemination level	PU			
DOCUMENT ID	D2.1 User needs and use cases			
FILE ID	BD2Decide D2.1			
Related documents	Technical Annex I DoA version 21/10/2015			

Distribution List

Organization	Name of recipients
AOP	T. Poli, E.M. Silini, S. Rossi, E. Martinelli, G. Chiari, C. Caminiti, G. Maglietta
VUMC	R. H. Brakenhoff, H. Berkhof, P. va de Ven
UDUS	K. Scheckenbach
INT	F. Favales, L. Licitra, E. Montini, G. Calareso, G. Gatta, A. Trama
UPM	G. Fico
Fraunhofer	F. Jung, S. Wesarg

[1] Status values: TOC, DRAFT, FINAL

Organization	Name of recipients
ATC	V. Tountopoulos
AII	A. Algom
POLIMI	L. Mainardi
MAASTRO	P. Lambin, F. Hoebers, A. Berlanga, G. Feliciani
MME	S. Copelli, F. Mercalli
UNIPR	M. Vitale, G. Gobbi, S. Rossi
European Commission	Project Officer: and all concerned E.C. appointed personnel and external experts

This document should be distributed as guidance to all the personnel of BD2Decide Consortium partners involved in the project execution.

Revision History

Revision no.	Date of Issue	Author(s)	Brief Description of Change
1	10.02.2016	E. Martinelli (AOP	ToC
1.1	15.02.2016	E. Martinelli (AOP)	Added contributions by UDUS and INT
1.2	24-03-2016	E. Martinelli (AOP)	Added contributions by INT, AOP and MAASTRO
1.3	07-04-2016	E. Martinelli (AOP)	Added new contributions by Licitra, Gatta, Trama, Berlanga
1.4	22-04-2016	E. Martinelli (AOP)	Added new user needs
1.5	27-04-2016	E. Martinelli (AOP)	Added section on population data (Italy)
2.0	30-04-2016	E. Martinelli (AOP)	Added contributions from INT, ISS and all technical partners
3.0	04-05-2016	E. Martinelli (AOP)	Added additional use cases for research, included comments received from partners

Addressees of this document

This document is addressed to the BD2Decide Consortium. It describes the users' needs and the operational environments where the BD2Decide DSS shall be used.

It includes the specifications of the data to be managed, the functionalities, operations, restrictions and mandatory requirements from the users side and from the privacy and security of data point of view, that are needed to orient the work of technical partners.

This document will be delivered to the European Commission.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

Abbreviations and definitions

CT (scans)	Computed Tomography scans
CT	Chemo-therapy
EHR(s)	Electronic Health Record(s)
HNC	Head and Neck Cancer
HNSCC	Head and Neck Squamous Cell Carcinoma
HIS	Hospital Information System
LIS	Laboratory Information System
MRI	Magnetic Resonance Imaging
OPSCC	Oro-Pharynx Squamous Cell Carcinoma
PET-CT	Positron Emission Tomography
QoL	Quality of Life
RT	Radio-therapy
RCT	Radio-Chemo-Therapy, i.e. treatment combining radiation and chemotherapy
TR	Tumor Registry

Executive summary

The objective of this deliverable is to present the user needs and requirements for the proposed technical and scientific infrastructure of BD2Decide. BD2Decide aims to deliver a state of art technological platform to facilitate collaborative clinical decision support and the provision of personalized care for cancer patients. The project will demonstrate this concept by means of a specific application to a particularly challenging clinical problem: personalized treatment decisions for late stage head and neck tumors.

The project's technical solution needs therefore to look ahead to emerging technologies for big data analysis and virtual representation of patients health data, that facilitate collaborative, objective, best informed and personalized decision-making.

The BD2Decide technological platform should therefore conform to the following main requirements and expectations of clinicians, in order to be accepted and adopted:

- **ubiquitous availability** that allows mobility (e.g. web-based applications, cloud-based data)
- **easiness of use**: support different languages, fast and reduced interaction times, adherence to clinical workflow, and fast response time (user-centred development, adaptive interactive user interfaces, patient-centred feedback mechanism enabling the clinician to distinguish the factors that had more influence for the prediction of a specific patient) ;
- **capability to process large amount of heterogeneous and complex** data and to produce valuable and reliable indicators to orient treatment (e.g. big data analysis, prognostic models);
- **collaborative environment**, to facilitate collaborative decision-making, improve treatment decisions, and foster collaborative research;
- **on-demand availability of software tools**: capability to provide the specific software for each situation as a commodity to support clinicians when required (Software as a Service, Service Oriented Architecture);
- **data privacy and security**: the BD2Decide architecture must allow the management of each patient inside each hospital (private/clinical environment) and the access to encoded/anonymized patient's data warehouses (shared/research environment) for the execution of the prognostic models.

The BD2Decide platform combines the complexity of the specific clinical domain to the technology challenges imposed by the need to align to technical innovations at the same time ensuring semantic and technical interoperability with legacy hospital information systems. Only if all these requirements are met the BD2Decide platform could successfully be adopted in clinical settings.

To address this complexity, user needs and use scenarios have been aligned according to two main "user needs and requirements" streamlines:

- **End user / clinical perspective**
- **Technical perspective.**

In this document we particularly focus on the end users and clinical perspective, however we provide a summary overview of the main technical needs that are mandatory for users acceptance and for the integration with legacy patient's data management systems (Hospital Information Systems, Laboratory Information Systems, Patients Health Records, Administrative Patients Data Systems). We also describe the use scenarios and the related needs for what concerns the inclusion of population data and of external knowledge (e.g. literature, PubMed, guidelines) and other external data accessible by BD2Decide Big Data System (e.g. public repositories of diagnostic images, public biomolecular datasets, tumor registries, epidemiological data repositories).

More details concerning technical needs and their implementation in the BD2Decide architecture will be provided in deliverable D2.3, when the final system architecture is defined.

About this document

This document is structured into six main sections and three annexes. The use cases and user needs are mostly concentrated in sections 1 through 3 (users perspective). Sections 4 through 6 provide additional information that impacts the technical development (technical perspective). The annexes include the list of data to be collected (CRF) and that will constitute the core of the BD2Decide Patients Documentation System (PDS). In details:

Section 1 provides an introduction to the BD2Decide concepts and usage in clinical practice and provides an overview of the clinical workflow.

Section 2 provides an overview of the users' needs, from the end-user perspective, but at the same time considering some legal and technical implications that must be taken into consideration by the technical developers.

Section 3 describes the use cases which will be used to guide the system development and will be used for the system validation. These use cases are examples of usual clinical activities and might be better detailed during the system development, as foreseen by the development approach that involves users with IT engineers.

Section 4 and 5 provide information concerning specific needs relevant for data security, data access, data preservation and privacy, data maintenance.

Section 6 is specifically addressing technical use cases.

The annexes contain information regarding the BD2Decide patient's data, population data and some information concerning international annotations (disease coding, disease localizations etc.).

1 INTRODUCTION TO USER REQUIREMENTS

1.1 THE CLINICAL CONTEXT OF BD2DECIDE

Head and Neck Cancer management is governed by international guidelines which consider the different localization and extension of tumors (TNM), such as NCCN[2] or ESMO[3] guidelines.

Therefore the clinical workflow and the patient's treatment approaches are currently standardized based on these criteria. With BD2Decide we intend to perfection such criteria and introduce new subcategories of patients characterized by different survival prognosis.

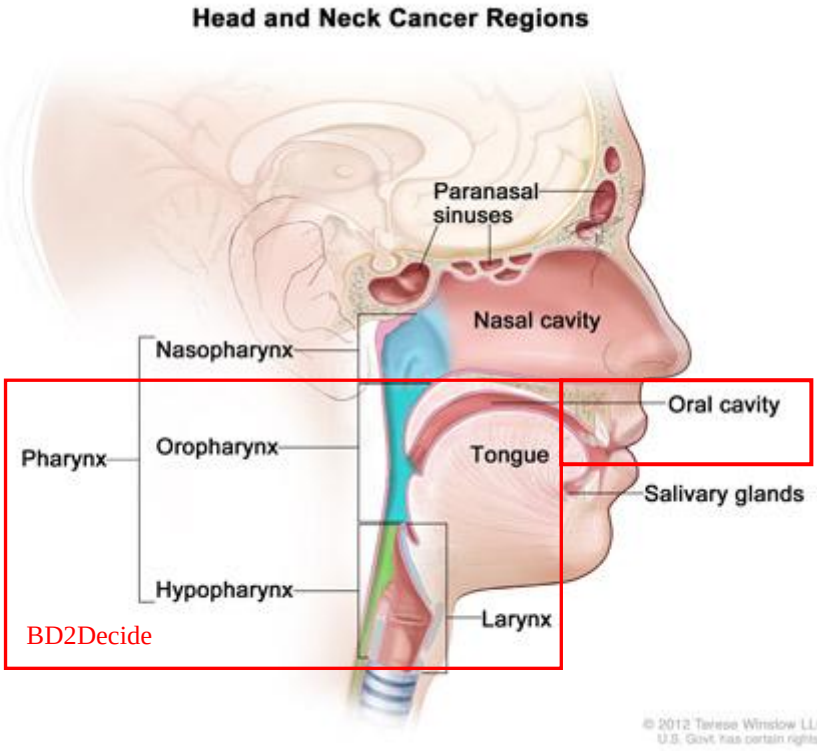


Figure 1. Head and Neck Cancer different localizations (in red squares the locations addressed by BD2Decide

Clinicians now must apply the specific guidelines to any patients with the same TNM and tumor localization and consider a few (less than five) additional clinical and pathological markers to decide the best treatment options. What users (physicians) need is a system able to produce reliable prognostications and related biomarkers and to stratify patients in subgroups by risk (High-medium- low-risk) with a higher confidence level than current criteria. So physicians can concentrate in the collection and analysis of the relevant prognostication factors and obtain decision support information personalized for each patient and which might be different in different populations (either from the geographic or the lifestyle and environmental point of view). This process should be available at the time of diagnosis, but might also be useful during the patient's

[2] http://oralcancerfoundation.org/treatment/pdf/head-and-neck.pdf

[3] http://www.esmo.org/Guidelines/Head-and-Neck-Cancers

follow-up, as a decision aid during adjuvant treatment and as an "alerting" in case prognosis deteriorates over time, to promptly diagnose possible reoccurrences. Thus a Clinical Decision Support alone is not sufficient: physicians need a transparent decision support aid, that allows their independent reasoning and that considers both evidence medicine (i.e. previous cases, literature etc.) and patient-specific prognostic factors and shows them clearly, along with their level of contribution to the prognosis prediction. In this document we intend to describe how users would like the new IT platform realized by BD2Decide to support them.

The clinical decision support system developed by BD2Decide integrates a number of specific software applications that should be used in daily clinical practice to collect and manage patient's data and to support the clinical workflow of care for such patients, from the first visit and the evaluation of patient's disease evolution, during the determination of the most adapted and effective therapeutic approach, throughout all the therapeutic path to the last follow-up. Additionally the platform should be linked to valuable external data resources, that provide population-related information which can determine important personalizations of the treatment decisions, such as lifestyle behaviours data or environmental exposures that could impact on the disease evolution.

As such the BD2Decide integrated platform must conform to the clinical workflows adopted in the different hospitals, to the specific use cases relevant for the different Head and Neck (HNC) cancers diagnosed and treated, and to the existing hospital information systems (HIS) components that could feed the BD2Decide patient's data repository.

Therefore it is paramount that BD2Decide conforms both to the needs of clinicians to manage large amounts of data with the minimum effort, at the same time ensuring that all relevant information for each patient is preserved and accessible, in a secure and privacy compliant way.

Additionally the involvement of the patients in the decision pathway is a core aspect addressed by the system, as patient's commitment and engagement in treatment has been demonstrated to be one of the most relevant drivers for disease recovery and quality of life improvement.

Because HNC treatment involves a multidisciplinary team of operators (surgeons, oncologists, radio- and chemo-therapists, biologists, radiology technicians, nursing personnel) and also the hospital administrative personnel, different levels of access and authorization must be foreseen as well as different views on the patient's data.

1.2 REQUIREMENTS ANALYSIS AND SCENARIOS

BD2Decide users are physicians and clinical researchers involved in Head and Neck Cancer (HNC) patients management and in clinical investigations. BD2Decide has the purpose to deliver a platform able to support both types of users: to support clinicians in personalized treatment decision making, to provide researchers with tools allowing the analysis of a huge amount of heterogeneous data for the identification of personalized prognostic signatures and for advancing knowledge on the biological and biomolecular aspects, presentations and characteristics of HNC.

The BD2decide platform architecture has therefore been conceived in two layers: a "private/clinical" layer, devoted and restricted to patient's management inside the hospital and a "research" layer, where anonymized data can be securely integrated and processed for new knowledge extraction.

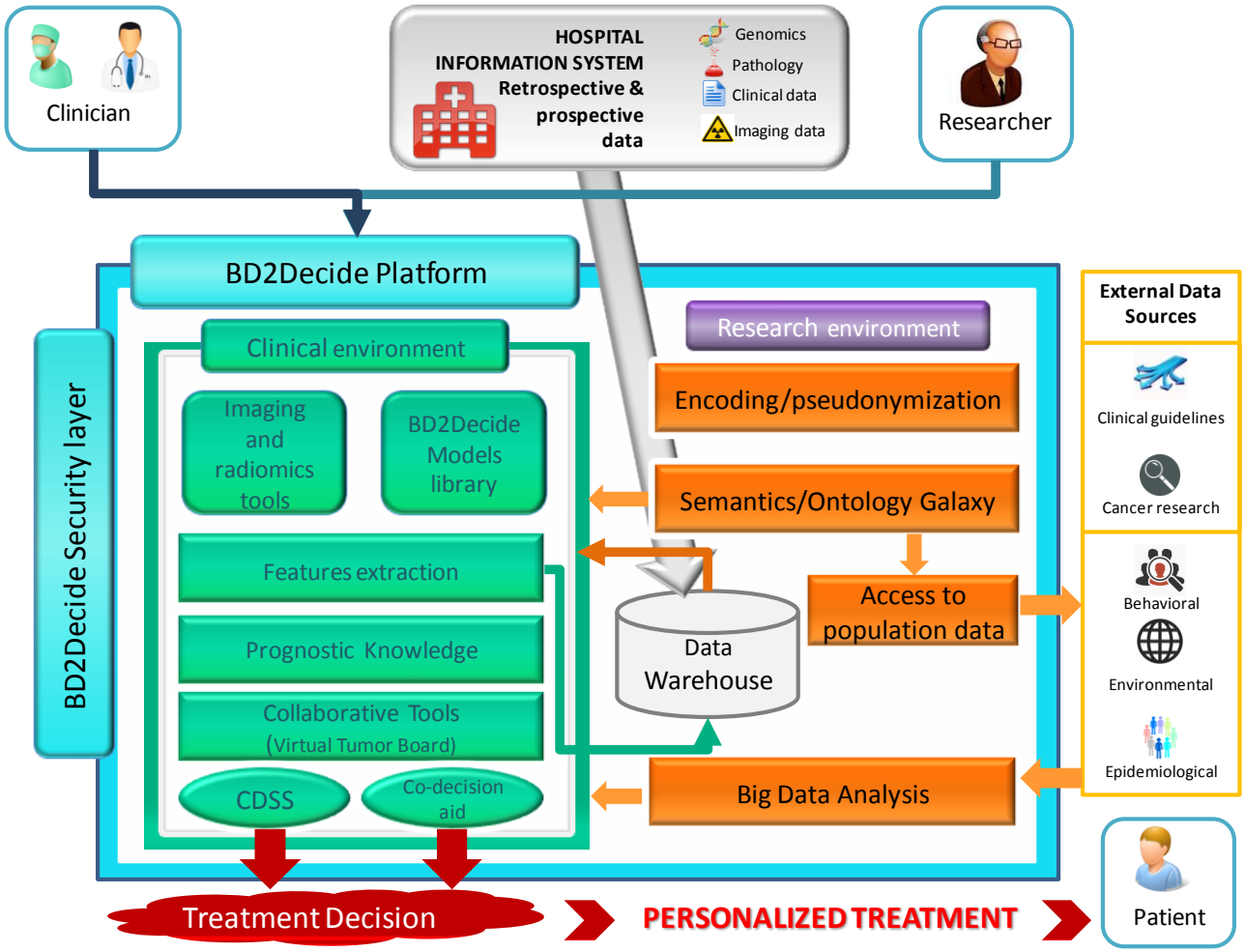


Figure 2. The BD2Decide architecture from users perspective

Clinicians will access the BD2Decide platform via a secure web access to use the BD2Decide tools to manage patients, execute the models and take collaborative treatment decisions. Users can also securely access the extended shared patients' data warehouse, under total privacy, to perform additional data analytics for research and knowledge sharing within the medical community.

Depending on the use scenarios, users might invoke the prognostic models from the BD2Decide CDSS which automatically gets the most appropriate model and the available patient's data, or may run the models from a web-based application.

Patients can participate to the final treatment decision via the BD2Decide co-decision aid, a web-based application that presents treatment options and empowers the patients by engaging them to a proactive role in treatment management.

BD2Decide tools, models and services might be used in research and for clinical decision support. Figure 2 above illustrates the two different use scenarios.

For research purposes patients' personal data are not needed, therefore the BD2Decide Consortium has agreed to share pseudonymized/encoded patients data in order to create a large database for HNC, accessible to the research community. To this aim a data transfer agreement has been conceived. By contrast, for clinical DSS patients' personal data will be needed: these data will not leave the hospital in charge and a secure layer will prevent any such data external exposure.

The visualization and representation of patients' data, the results of the CDSS are needed in a short timeframe, to allow physicians take the best treatment decisions in the shortest time. This implies a very effective workflow management for the collection of all the required multiscale data (clinical, imaging, biomolecular and genomic, etc.), for data sharing and the execution of the prognostic models and for the shared decision-making. Form the IT perspective this implies the availability and adoption of high-performance computing able to access data in different/scattered data sources (Big-Data, Cloud computing, cloud-based systems).

A system administration/authentication service is needed to manage users' access rights, and to regulate which data and tools shall be made available to each user.

In this scenario the possibility to exchange data with other health care systems is of paramount importance, thus the BD2Decide platform shall interoperate with other systems throughout the patient's disease and health management system. This need is obvious when we consider already in place Hospital Information Systems applications, but is also extremely relevant for the overall patient's health records. To prepare this integration, which is not yet achieved by the majority of EU hospitals, the most accredited standards shall be adopted to support communications, messaging, encoding standards to communicate with HIS and EHR systems.

Give the high incidence of adverse outcomes for these tumors, the speed of intervention is crucial, therefore the care delivery workflow, the diagnostic operations and the data collection and analysis timeline must be synchronized to achieve the shortest elapsed time between diagnosis and primary treatment (see Figure 3).

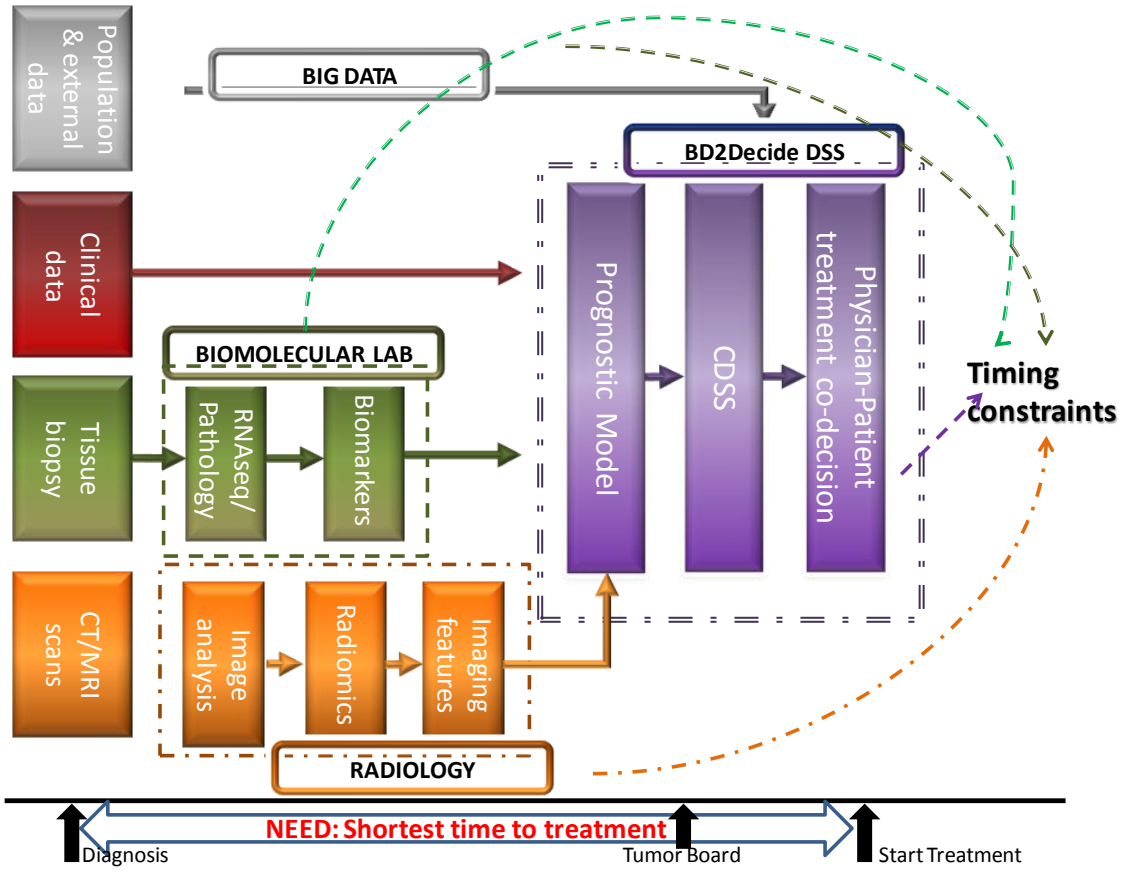


Figure 3. BD2Decide workflow and need to fast decision-making (CDSS) for HNC patients

1.3 THE CARE DELIVERY WORKFLOW

HNC patients management workflow usually consists in three different phases:

1. diagnosis: this phase starts when the patient is referred to the hospital for verification of possible Head and Neck Cancer and ends at the moment when the patient treatment is determined (surgery or other therapy).

2. treatment (and hospital care when needed): this phase starts when the patient either (i) enters the hospital for surgery or (ii) follows the non-surgical treatment pathway, and ends (i) at the moment of discharge from hospital or (ii) at the end of the treatment cycle.

3. follow-up: this phase starts usually one month after the end of treatment, and lasts up to 5 years with different frequency of follow-up visits depending on the tumor type, on the applied treatment and on the patient's status.

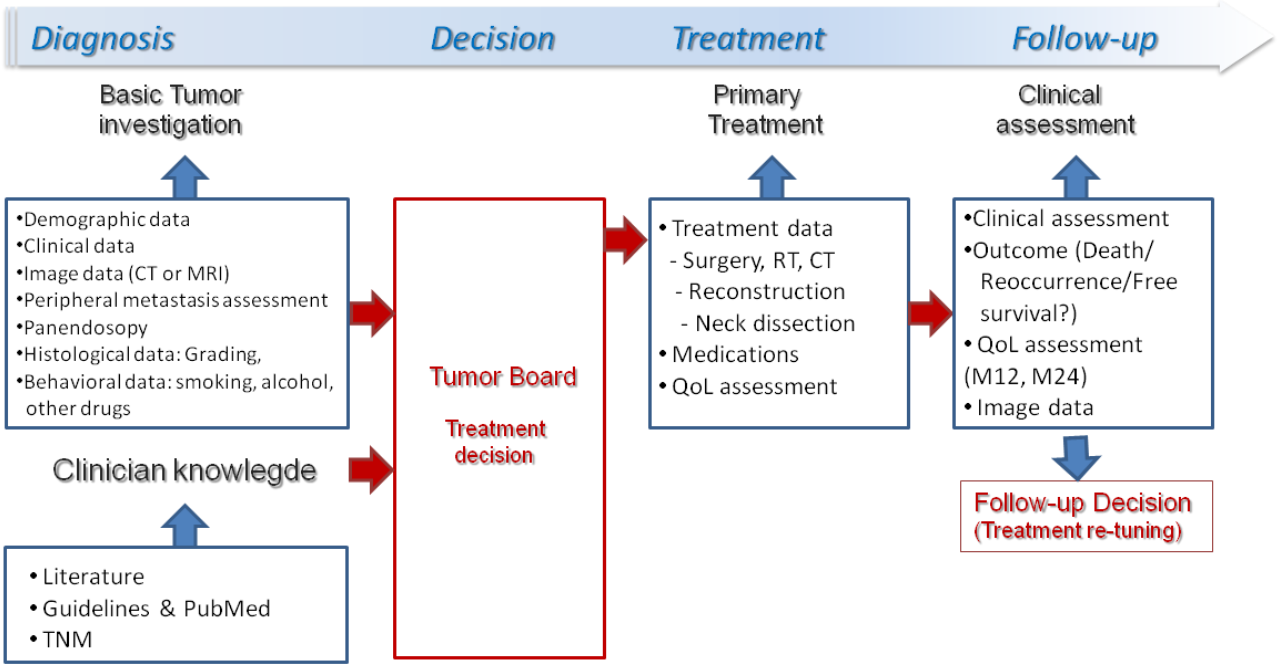


Figure 4 Current workflow

1.3.1 Current workflow

1. Diagnosis.

The patient is examined by the physicians (surgeons, oncologists) after presenting some signs and symptoms. In this phase the physician performs a visual examination of the patient and some biopsies of the affected regions, in order to determine the onset and the extent of the disease (clinical staging). In case the patient already has a result of biopsy, the surgeon evaluates the results. Additional exams are then conducted (pathological, radiological, etc.) that provide a more precise diagnosis and a further visit is planned. To exclude a secondary primary and to better assess the extension of tumor growth, a panendoscopy is performed in most cases. The patient is inserted in

the hospital records, with the relevant data (EHR).After the results of the biopsy are available and the diagnosis is histologically confirmed, the patient is classified for HNC treatment.

Data from the clinical exam are collected directly by the physician during the visit, pathology data may be directly collected from the Laboratory Information Systems when available in the hospital, radiology data (CT/MRI scans and referrals) may either be brought by the patient or retrieved from the hospital PACS. In the first case the electronic format of the radiological exams may be either copied to be kept in the patient's records and be imported into the PACS system of the clinic.

A preliminary case discussion takes place (pre-treatment Tumor Board). In alternative the surgeon/oncologist consults the pathologists and - more often- the radiologist by phone. At the end of this phase the patient situation is evaluated by a multidisciplinary team (Tumor Board) and the best treatment approach decided. The patient is informed by the nursing personnel and/or is recalled for a talk with the physician in charge and is enlisted for treatment.

In this phase, the participation of the patient to the treatment decision is crucial, therefore the physician explains to the patient the situation, describes the treatment alternatives if available, indication pros and cons. A consent to treatment is then collected.

For the specific BD2Decide clinical study, advanced stage (Stage III, IV) HNC patients complying with inclusion and exclusion criteria may be enrolled, upon informed consent.

1.1 Decision making process

Decisions are taken on two occasions: during the multi disciplinary visits (the patient is present) or during the weekly tumor board were the patient case is discussed.

The multidisciplinary team usually involves: surgeons, radiation oncologists, clinical oncologists, radiologists, pathologists, chemotherapists and nursing personnel.

1.1.1 Decision criteria

After clinical diagnosis the physicians involved study the patient before confronting and taking decisions. The main criteria for treatment decisions are based on the TNM staging system and on the gold standard clinical practice and international guidelines (e.g. NCCN guidelines[4]). These guidelines are mostly relying on the TNM staging and on a few additional conditions (e.g. positive margins, extracapsular spread, multiple positive nodes, perineural/lymphatic/vascular invasion etc.)

1.1.2 Treatment decision

In case of multidisciplinary visits, a single patient is discussed. In case of tumor board (usually held on a weekly basis) many patients are discussed by the multidisciplinary team.

Decisions may be taken also by involving some physician of the multidisciplinary team by phone, when their presence is not possible. In some hospitals this procedure is performed also for single cases, asynchronously during the week or in case a decision must be taken on emergency.

[4] see: http://www.nccn.org/professionals/physician_gls/f_guidelines.asp#site

Given the limited time for tumor board / multidisciplinary visits, the most relevant information concerning the patient/disease must be available.

The shared decision is recorded and signed by the chair of the tumor board or of the multidisciplinary visit. Decisions are taken before the first (primary) treatment and also during follow-up visits, as a consequence of the curative (non)effectiveness of the treatment.

1.2 Involving patient in treatment decisions

In case of multidisciplinary visits the patient is contextually informed concerning the treatment options and can take informed decisions jointly with the physicians.

In case of treatment decisions taken during Tumor Boards, the patient is informed by the caring physician during a pre-treatment/treatment assignment visit. In this case the physician presents the situation, the treatment options and the proposed treatment decision taken by the multidisciplinary team. The patient and the physician take agreements concerning the treatment to be administered.

Quality of life of the patient might be assessed before treatment (optional) and the expectations regarding future quality of life are also evaluated.

2. Treatment

Depending on the extent and localization of the tumor, different treatments may be chosen and applied: surgery eventually followed by radio- chemo-therapy, radio- or chemo-therapy alone or the combination of both. Also the type of drug and the dosage are determined.

Dependent on whether surgery is performed or not and on the patient´s performance, different dosages and radiations regimens are applied. Chemotherapy is mostly platinum-based, but may also include 5-FU, Paclitaxel, EGFR (Epidermal Growth Receptor Antibody) and others.

Treatment data are collected during surgery when applicable (surgical procedures applied, adverse events etc.) and in case of non-surgical treatment (radiotherapy, chemotherapy, or combined RCT). Treatment data are automatically recorded in the patient's EHR by means of the Hospital Information System specific applications.

In case of surgery, post treatment data are also recorded including estimation of quality of life at discharge.

In case of RT, CT or combined RCT, at each scheduled and unscheduled visit data concerning the patient status are recorded.

3. Follow-up

Follow up visits and examinations are usually performed to evaluate the effectiveness of the delivered treatment by checking the tumor status in term of remission, to support cured patients to overcome late toxicities induced by the treatment itself, and possibly for tumor recurrence anticipation or early discovery of a second primary tumor with ultimate the aim to facilitate salvage

treatment. However this last point is still very controversial since at present there is no clear evidence that an intensive visits and examination scheduling is effectively impacting on survival. In this scenario follow up modalities as well as timing that are adopted in different Centers may differ. Our project includes very experienced Institutions so that we are confident that the patients treated in this Hospitals will be treated and followed according to the general principles inspiring the global care of the individual patient. In general visits and imaging are done more frequently in the first 2 year (q 3 months to q 6 months after the end of therapy) and tends progressively to diminish till 1 per year. In general with 5 year the follow up procedures are closed.

During follow-up period the patient is visited by the physician in charge every 3 months at least for the first two years, then every 6 months and after year 3 once a year. During each follow-up visit the following data are recorded: status of patient, remission, reoccurrence, other signs / symptoms. In case of need new diagnostic images may be collected. This is normal for OSCC patients, who are required to perform a CT/MRI every 6 months during the first 2 years of follow-up. In this case diagnostic images are recorded in the hospital PACS and can be retrieved by clinicians.

For patients undergoing radio/chemotherapy, the overall status is assessed and - in case of side effects, toxicity issues or non effective treatment, a new treatment may be decided. In this case a consultation within the multidisciplinary team may be performed.

Quality of life is also re-assessed at each follow-up visit (interview with the patient).

In case a patient does not show-up, the relevant information (patient lost during follow-up) might be recorded. In case the patient is dead, the information whether the patient is dead for the disease or for any other cause is also recorded.

1.3.2 Workflow with BD2Decide

BD2Decide introduces new knowledge and new supporting evidence and tools in the decision-making process. This implies the collection of new biomolecular data (e.g. genomic and HPV) and the identification of specific tumor phenotypes through the processing of diagnostic images. These operations require specific tools and new workflow operations, which need to e carefully synchronized to prevent extending the timeframe between diagnosis and treatment. BD2Decide is expected to provide all these functionalities within the current workflow timeframe, and even to speed-up the decision process.

1. Diagnosis

The BD2Decide Software will accompany the while patient´s workflow, as soon as a head and neck tumor is suspected. When the patient presents the first time, his demographic data, risk factors and anamnesis are inserted. As soon as investigations lead to further information like pathology after the biopsy and imaging as well as further results by performing a panendoscopy, also these data in fed into BD2Decide. To handle imaging data in an easier way, automatic segmentation and characterization is performed using the Fraunhofer tool. This is also the basis for radiomics, which are performed after automated segmentation on the pictures. The pictures are saved in hospital

PACS and the segmented images should be saved after pseudonymization/encoding into the BD2Decide patient's data repository (Patients Documentation System), The features resulting from diagnostic images processing (image analysis, radiomics on CT and on MRI) are integrated into the BD2Decide data repository and can be retrieved and visualized by clinicians (all involved clinicians). The radiologists might include selected images in the EHR dataset that better represent the patient's tumor and lymph-nodes, for immediate visualization by oncologists, surgeons, radio- and chemotherapists.

As soon as the biopsy leads to the result of a squamous cell carcinoma, pathologists are informed to do a microdissection of representing tumor specimens to pass it to the genetic department. Here, genomic investigations are performed using the BD2Decide biomarker profile to further characterize the tumor on a molecular level. Results are as well integrated into the BD2Decide Patients Documentation System and are also recorded in the EHR for the patient inside the hospital.

1.1 Decision making process

1.1.1 Decision criteria

After these steps, the BD2Decide software integrated clinical, radiological (including radiomics) and genetic (own biomarker profile) data of each individual patient. Using the underlying epidemiologic and integrated data with the BD2Decide algorithms, BD2Decide is now able to show a virtual prognosis of the patient.

At the time of diagnosis and when all clinical, imaging and biomolecular data (genomics, histology etc. from biopsies) are available, any member of the multidisciplinary team involved in patient's treatment can run the prognostic model and assess the patient's survival probability, and the most relevant factors responsible for the predicted outcome. The BD2Decide platform will propose the most appropriate prognostic model and present the prognostic prediction both graphically (survival curves at 1, 2 and 5 years, the positioning of the individual patient with respect to the population of patients with the same tumor stage/characteristics) and with analytical information (i.e. list of prognostic factors with their value, relevance, confidence interval for the prediction).

The physicians will be able to invoke different models by selecting them from the models library, in order to compare results and obtain additional evidence. The clinician can also visualize the most relevant information as defined by current guidelines and so achieve a full background of information for treatment decision.

The special aspect of this forecast is, that here not only TNM but many other impairing factors are respected in a way even experienced medical specialists could never perform. The BD2Device can now be used as objective tool for discussion of the patient´s treatment within the tumor board and there visualize the obtained results in a compact, easily understandable, partly graphical way.

1.1.2 Treatment decision

Decisions may be taken by the multidisciplinary team either as usual (physical meetings) as defined in section 1.3.1 or by means of web-supported Tumor Board. BD2Decide offers a web-based tool (Virtual Tumor Board) that allows the physician in charge of an individual patient to plan a Tumor

Board Meeting and invite the members of the multidisciplinary team. The web-based tool also supports physical meetings, as it records the discussions and the decisions taken for each presented patient's case. The web-based tool allows shared display (screen sharing option) of the relevant patient's data (including diagnostic images etc.) when more than one member of the multidisciplinary team is present. In case a member of the multidisciplinary team is requested a on-the-spot consultation, he/she can remotely access the BD2Decide platform, recall and visualize the patient's data and the patient's case presentation if already recorder by the colleagues requiring the consultation, and can add his/her recommendations and comments. These actions can be performed both synchronously (i.e. at the same time as other members of the multidisciplinary team are connected) or asynchronously (i.e. anytime before the treatment decision).

The physician responsible for the patient (or the chair of the Tumor Board) can invite external experts for side consultations, by means of the BD2Decide Virtual Tumor Board tool. This functionality may prove particularly useful for minor hospitals that are linked to main Cancer Centres, or when a particularly difficult case is under discussion.

All discussions are recorded and digitally signed by the clinicians and a final consensus report for each case is recorded and printed if needed.

It is expected that by introducing BD2Decide tools clinicians and patients will be enabled to take more personalized decisions. This will be possible through the integration of many variables that at present cannot be considered due to their number and interaction complexity.

The new tool will be able to deal with them and will render a very punctual information that is what is needed in individual decision making. So for example we will be able to precisely foresee single patient outcome under different treatment scenarios. This will have an impact on clinical management and quality of life of individual patient. It will have to be demonstrated whether it will also have an impact on patient population outcome, with respect to classical decision making. Ideally this will have to be performed with a randomized study where subjects will be treated with the aid of the decision tool or without and their outcome will be compared. Our guess is that this kind of approach will be considered obsolete by the time big data will be routinely used and fully exploited.

1.2 Involving patient in treatment decisions

Web-based co-decision tools developed by BD2Decide will be available to inform the patient, after the diagnosis. These tools will be specific for each Head and Neck tumor subsite and will present some information concerning the specific tumor, the different treatment options and the pros/cons of each treatment option. These tools should be available also on tablet PCs (or smartphones) and include interactivity possibilities, with questions related to the side effects and expected outcomes probabilities of the possible treatment options. The patient may view and use these tools independetly from the physician before the pre-treatment decision visit, in order to get independent information, or can be used at the presence of the physician. In this latter case the physician can complement the co-decision aid tool with more specific answers to questions and doubts from the

patient. An example of such tool has already been realized by MAASTRO for prostate cancer (see http://treatmentchoice.weebly.com/prostate-cancer.html).

The tool can be linked to the patient's prognostic model results and thus be more precise concerning the survival probability (personalized prognostic prediction).

The patient's preferences are recorded (patient's profile collected by the decision aid tool) and can be visualized by the physician before the visit with the patient, in order to better orient the discussion between the physician and the patient.

The final treatment decision is taken by the physician in agreement with the patient.

2. Treatment

Treatment is administered as usual practice (see above 1.3.1). The main advantage of BD2Decide is the possibility to visualize all patient's data and the prognostic prediction anytime during the treatment. Treatment data are collected either through the BD2Decide interfaces or directly from existing Hospital Information System application. To achieve this, integration with existing HIS will be implemented (or prepared by means of standards for data exchange).

3. Follow-up

We expect that by implementing BD2Decide tools the clinicians will have the opportunity to personalize the follow up schedule based on the foreseen individual outcome. With a better assessment of the potential patient´s outcome, it will be easier to personalize his/her follow-up. The frequency of personally showing up in the clinic for follow-up examinations can be adapted (more for risky patients, less for prognostically “good” cases), the same may be performed for imaging. Therefore follow-up is performed more detailed for those who need it and more comfortable for those, who have the luck to be at low risk. The frequency of imaging has also an impact of the radiologic contamination of each patient, at least when CT scans are necessary. This will reduce unnecessary costs and will concentrate resources where more likely needed.

During follow-up the status of the patient is recorded and the quality of life is re-assessed periodically as foreseen by the clinical protocol, though questionnaires. The improvement in quality of life is measured by means of data analysis.

2 BD2DECIDE USERS AND USERS' NEEDS

2.1 ACTORS INVOLVED IN HNC TREATMENT

Actors are the persons concerned with HNC patients management and who will use the BD2Decide platform for patients' data collection and analysis and for decision making, based on BD2Decide prognostic models personalized predictions. Actors are therefore the healthcare professionals in charge of patient's management and the patients themselves. These actors are usually supported by existing software modules from the Hospital Information System (HIS) in some operations of the care delivery workflow, where significant data are collected (e.g. laboratory information systems, treatment recording software platforms, radiological exams evaluation platforms), which must be considered in order to determine restrictions and standards for software and data integration.

BD2Decide usage and with the results of BD2Decide models prediction, i.e. all the clinicians involved in the OSCC patients management and the patients themselves. Other indirectly concerned actors include in particular the IT Department of the hospital and the hospital administration. To these we should add other external actors providing epidemiological data at the level of populations or even at the level of the single individual (e.g. population-based cancer registries, territorial health services collecting medication data).

We describe in the following the roles and access rights of each actor.

2.1.1 Health professionals

Surgeon, Medical Oncologist, Radiation Oncologist

The surgeon, the radiation oncologist and the medical oncologist are the main users of the BD2Decide system as they are the clinicians responsible for the diagnosis, treatment and follow-up of the patient.

These professionals use the BD2Decide system for the following activities:

- Check patient's data and visits
- Organize and check daily agenda for visits, surgery
- Plan and record visits and related data
- Plan and record hospitalization when needed
- Plan and record surgery when needed
- Record hospital care procedures and patient's data
- Participate/Record Tumor Board decisions
- Communicate with patient and involve the patient in treatment decisions
- Plan and record radio- and chemo-therapy when needed
- Plan and record follow-up visits and relevant data
- View all patients' data including imaging

- Run risk prediction model and perform simulations
- Check patient's quality of life

The surgeon is authorized to insert, modify and delete (upon checked confirmation) clinical data, treatment and follow-up data for the patient, view all data and to visualize data from pathologists and radiologists and the data inserted by the Medical Oncologist and the Radiation Oncologist and of course all the data automatically transferred from the HIS.

The medical oncologist and the radiation oncologist are authorized to enter patient's clinical data and to manage all data concerning radio and chemotherapy and follow-up. They have the right to visualize all the remaining data.

Radiologist

- Perform CT/MRI and produce referrals or receive CT/MRI from external
- Anonymize imaging data and transfer to Image Analysis Tools
- Analysis and post-processing of image datasets and data recording of results related to imaging
- Extract radiomics features
- Manage imaging and radiomics data
- View all patients' data
- Participate to Tumor Board.

The radiologist is authorized to enter, delete and modify the data related to radiology exams and has access to view only the remaining data of the patient.

Pathologist

- Perform histo-pathological exams and record data (or transfer data from existing information systems)
- Perform immuno-histo-chemical exams and record data (or transfer data from existing information systems)
- Participate to Tumor Board
- Monitor biological samples to be analyzed
- View all patient's data.

The pathologist is authorized to enter, delete and modify the data related to pathology exams (histology, immunologic tests, etc.) and has access to view only the remaining data of the patient.

Biomolecular analyst

- Perform RNA extraction, Genomic data analysis and upload
- Visualize patients to be processed

The biomolecular analyst has the right to upload / delete / download the genomic data collected. He/she has the right to visualize all the remaining data.

Nursing personnel

- Perform pre-hospitalization take in charge of patient
- Perform hospitalization of patient (i.e. accept patient for surgery)
- Perform patient care during hospitalization and view prescriptions for the patient on EHR
- Call patients for radio-chemotherapy visits
- Create patient hospitalization record.
- Verify the Electronic Health Record (EHR) to check exams done to the patient.

Nursing personnel have limited access, only to visualize patient's data related to hospital care and discharge. The nursing personnel may activate some functionality such as: printing the discharge letter or specific referrals/reports to be signed by the surgeon, activate booking of follow-up visits and of other exams such as radiology exams (CT/MRI). To manage authorization for nursing personnel, data sub-sets must be defined and functionalities defined which may be used by nursing personnel. Nursing personnel in Amsterdam also participate to Tumor Board meetings.

Clinical researcher

The Clinical Researcher has only access to data for research purposes, i.e. has only read rights to patient's full records in the cases foreseen by the relevant hospital authorization system and to anonymized data and to the DSS.

2.1.2 Hospital IT personnel

- Manage the OraMod platform.
- Ensure data integrity, security and recovery.
- Ensure interoperability (whenever possible).

The hospital IT personnel is in charge of the maintenance of the OraMod platform and of the supervision of all actions intended to interoperability with existing HIS software. One system administrator will be identified and appointed at the end of the project, to maintain the OraMod platform.

2.1.3 Hospital Administrative and Legal Offices

- Manage patient's demographic, administrative and financial data.
- Issue administrative and clinical documents (including patient's records and surgery data) to the patient.

- Manage costs.
- Guarantee patient's privacy.
- Guarantee data preservation.

The Administrative personnel of the hospital is not directly involved in the management of BD2Decide data but should be involved anytime data from BD2Decide must be used for administrative purposes. These persons do not have access to the BD2Decide platform, unless invited with visualization rights only.

2.1.4 Hospital Departments specifically interested in BD2Decide

- Maxillo-facial surgery
- ENT
- Oncology
- Radiation oncology
- Pathology
- Radiology
- Biomolecular laboratory
- Administration and patient's acceptance
- Emergency unit.

2.1.5 External entities potentially interested in BD2Decide

- Cancer registry
- Territorial health services.

BD2Decide should automatically generate data for the cancer registry active in the hospital regional/national network. Cancer registry should have access to the patient demographic data, to the pathological information, to the biomarker profile, and to the patient's data related to hospital care, discharge and follow-up visits. This functionality will be eventually assessed during the technology development. The system should also be able to collect data from external data sources and use them for better prognostication.

2.2 HOSPITAL INFORMATION SYSTEMS COMPONENTS RELATED TO BD2DECIDE

- Management of visits (outpatients visits)
- Pre-hospitalization software (take in charge of patients): manages the waiting-list for hospital treatment.

- Emergency department software for patient acceptance
- List of visits (radiology)
- Surgery procedures recording (registry) and relevant safety (infections control, etc.)
- Patient's demographic data management (administration)
- Management of prescriptions
- Management of radio and chemotherapy
- PACS and any imaging software
- Visits reservation
- Request for histology exam
- Laboratory Information System for lab exams and for histology.

2.3 LINKING BD2DECIDE DATA TO THE PATIENT

A secure system to link the hospital patient's data to BD2Decide anonymized data shall be implemented in each participating hospital. The information linking personal data to anonymized patient's data shall be kept inside the hospital secure intranet.

3 USE CASES

This chapter illustrates how users (actors) will use the BD2Decide platform during their daily practice, according to the workflow presented in chapter 1.3.2.

General requirements

The following general user requirements shall be met, to ensure data quality and avoid errors:

1. Integrate BD2Decide with existing HIS as much as possible.
2. In case of input from clinical devices or paper-based referrals the use of automatic patient's code readers (e.g. bar code readers) shall be implemented to ensure that the data are correctly assigned to the right patient.
3. Design the users interactions in order to minimize the unnecessary data input and the number of "mouse clicks" needed to record or to visualize data.
4. Allow data entry/data retrieval on portable devices to facilitate access.
5. Implement invitations to Tumor Board, alerts concerning patient's critical situations both on BD2Decide platform and, when applicable, also on the physician's mobile phone, using

3.1 DIAGNOSIS OF PATIENT

First visit

Use case. A patient of 65 year in general good health condition presents to the general practitioner with a swelling in left side of the neck. The patient is referred to a general oral and maxillofacial surgeon or otolaryngologist, or directly to a head and neck cancer center.

The patient is always referred to one of the head and neck cancer centers, and an appointment is made at the outpatient center of in this case the VU University Medical Center.

User needs. *The clinician enters BD2Decide platform and opens a new EHR for the patient, using the patient ID provided by the hospital and stamped on the visit prescription issued by the hospital at the time of visit reservation/payment. BD2Decide visualizes demographic data of the patient if already registered in the hospital or allows entering basic demographic data. The new folder (EHR) is created.*

All demographic and clinical data in the medical system should be integrated into the software automatically.
When the patient is first diagnosed with oral cancer, there are only few information available. It would facilitate the clinician´s work, if he could sketch the tumor localization and dimensions on a anatomical drawing (see 2.3.1), showing all sites of the Head and Neck region affected. (Example: oral cavity See Annex I and Figure 5)
The marked anatomical sub-localizations should be automatically identified and translated into the data evaluation of BD2Decide, e.g. if the clinician marks the whole left tongue on the sketch, the software would

transfer the information, that the tumor is in the tongue on the left side into the data set. To standardize the registration of head and neck cancer sites, it could be very useful to provide the full description of the site of the tumors together with the corresponding ICD10 and ICD-O3 which is the standard classification used by cancer registries.
Furthermore it should be possible to integrate selected images from CT/MRI (selected by radiologists) or pictures schemes of the tumor into the system and possibly synchronize them with the anatomical sketches or/and also mark the tumor in the photography.

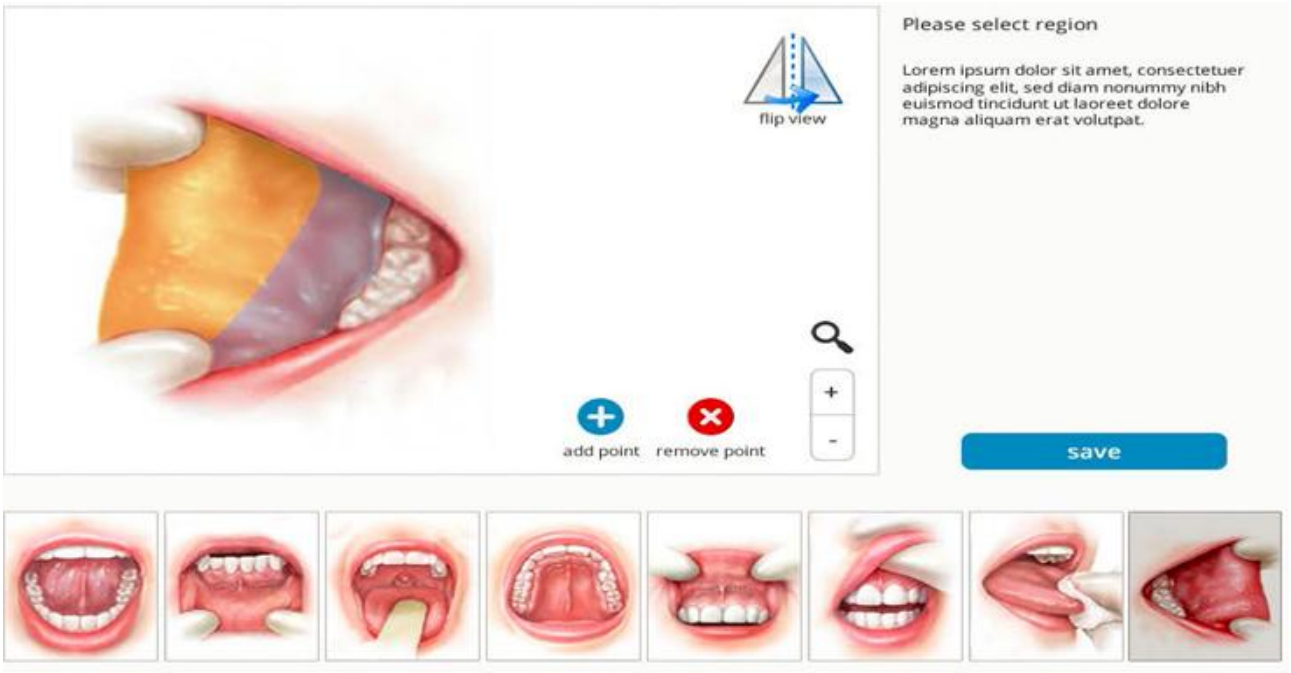


Figure 5. Drawing the tumor localization during a patient's visit (Oral Cancer). Image to be saved in BD2Decide data base. Similar drawings will be used for different localizations

Head and Neck Tumors localization (T-localization) relevant for BD2Decide, according to ICD10 international coding.

Oral cavity: C02.0-C02.3, C02.9, C03.0-C05.0, C06.0-C06.9; LarynxC32; Oropharynx C01.9, C02.4, C02.8, C05.1-C05.2, C05.8-C05.9, C09.0-C10.3, C10.8-10.9, C14.2; Hypopharynx C12, C13.

Cancer of the lip (ICD-10: C00), tongue (base of tongue: ICD-10: C01; other/unspecified part of tongue: ICD-10: C02, and other parts of the oral cavity gum: ICD-10: C03; floor of mouth: ICD-10: C04; palate: ICD-10: C05; other/unspecified part of the mouth: ICD-10: C06.

Malignant neoplasm of oropharynx, unspecified (ICD-10:CM C10.9), Malignant neoplasm of overlapping sites of oropharynx (ICD-10-CM C10.8), Malignant neoplasm of lateral wall of oropharynx (ICD-10-CM C10.2), Malignant neoplasm of posterior wall of oropharynx (ICD-10-CM C10.3), Malignant neoplasm of oropharynx (ICD-10-CM C10)

Malignant neoplasm of hypopharynx (ICD-10-CM C13), Malignant neoplasm of hypopharynx, unspecified (ICD-10-CM C13.9), Malignant neoplasm of overlapping sites of hypopharynx (ICD-10-CM C13.8), Malignant neoplasm of posterior wall of hypopharynx (ICD-10-CM C13.2)

Malignant neoplasm of pharynx, unspecified (ICD-10-CM C14.9), Malignant neoplasm of overlapping sites of lip, oral cavity and pharynx (ICD-10-CM C14.8)

Malignant neoplasm of larynx (ICD-10-CM C32), Malignant neoplasm of larynx, unspecified (ICD-10-CM C32.9), Malignant neoplasm of overlapping sites of larynx (ICD-10-CM C32.8).

Ref. http://apps.who.int/classifications/icd10/browse/2010/en#/C00-C14.
For the identification of the loci in anatomic images please refer to Annex I

Head and Neck Tumors localization (T-localization) relevant for BD2Decide, according to ICD-O3 international coding.

8070/2 Squamous cell carcinoma in situ, NOS; Epidermoid carcinoma in situ, NOS; Intraepithelial squamous cell carcinoma; Intraepidermal carcinoma, NOS.

8070/3 Squamous cell carcinoma, NOS; Epidermoid carcinoma, NOS; Squamous carcinoma; Squamous cell epithelioma.

8070/6 Squamous cell carcinoma, metastatic, NOS.

8071/3 Squamous cell carcinoma, keratinizing, NOS; Squamous cell carcinoma, large cell, keratinizing; Epidermoid carcinoma, keratinizing

8072/3 Squamous cell carcinoma, large cell, nonkeratinizing, NOS; Squamous cell carcinoma, nonkeratinizing, NOS; Epidermoid carcinoma, large cell, nonkeratinizing.

8073/3 Squamous cell carcinoma, small cell, nonkeratinizing; Epidermoid carcinoma, small cell, nonkeratinizing.

8074/3 Squamous cell carcinoma, spindle cell; Epidermoid carcinoma, spindle cell; Squamous cell carcinoma, sarcomatoid.

8075/3 Squamous cell carcinoma, adenoid; Squamous cell carcinoma, pseudoglandular; Squamous cell carcinoma, acantholytic.

8076/2 Squamous cell carcinoma in situ with questionable stromal invasion; Epidermoid carcinoma in situ with questionable stromal invasion.

8076/3 Squamous cell carcinoma, microinvasive

Patient examination and results

Use case 1. The patient is visually inspected and all relevant **clinical, socio-demographic,** and behavioral information collected. Visible suspect areas may be **biopsied** for histological examination. Otherwise, when a swelling of the neck is detected, cytological aspiration under ultrasound guidance will be performed by the radiologist. Therefore, imaging **by CT or MRI** will be acquired on indication. In this case, a suspect lesion on the left side of the tongue is seen and, therefore, biopsied by the surgeon.

User needs. *BD2Decide must collect and present clinical,socio-demographi, and behavioral data,*

biopsy results, CT and/or MRI scans. Patient's demographic data (sensitive data) are just shown (retrieved from HIS) but not recorded in BD2Decide database. Data related to physical exam done by the surgeon are manually inserted. Data from histology are collected from Laboratory Information System (if possible) or by manual input (alternative solution) by the pathologist. The surgeon classifies the tumor as Stage III (advanced stage), to be further confirmed by biopsies.

If imaging is requested, the paper request for imaging exam must be issued through BD2Decide and if possible (i.e. if a HIS is in place) the imaging exam will be booked for the patient. As soon as the imaging exam is analyzed by the radiologist and processed by the BD2Decide Image Analysis tools, the data from imaging must be recorded and be available to the surgeon (on BD2Decide system and on hospital PACS) and will be used for radiomics features extraction. In case imaging data is performed outside the hospital, a CD with the imaging dataset must be provided by the patient at the next visit, copied and sent to the radiologist for Image Analysis.

Use case 2. A 55 year old patient is referred to otolaryngologist by her general practitioner after a 2 weeks course of medical therapy for dysphonia without benefit. Standard workup demonstrate a larynx mass involving glottic region. **Biopsy** for histological examination confirm the suspected diagnosis of a squamocellular carcinoma; imaging **by CT or MRI** are acquired as requested.

***User needs**. The surgeon classifies the tumor as Stage III (advanced stage), specifically a glottic cT3 N0 M0 with invasion of the paragloctic space. Several treatment options are available for such cases (i.e. total laringectomy, induction chemotherapy followed by radiotherapy, concomitant chemo-radiotherapy). These approach present very different outcomes in terms of impact on quality of life (QoL, i.e. ability to speak, quality of voice, swallowing, etc…).*

BD2Decide, collecting clinical and radiological information, gives the clinician a feedback regarding patient's prognosis and QoL data based on stage and standard treatment, comparing the different treatment option. The collected patient's data shall be confirmed (signed) by the physician.

Use case 3. A 57 year old man is referred to otolaryngologist by anesthesiologist because of tonsillar swelling occasionally detected during routine workup for general anesthesia planned for correction of inguinal hernia.

Biopsy of the left tonsil shows spinocellular carcinoma with p16 positive staining. The lesion seems quite small, apparently limited to the tonsil without clear lymphatic spread. Node status investigated by MRI and FDG PET scan remains doubt due to borderline enlargement of para-pharyngeal lymph node .

***User needs**. The surgeon, the radiation oncologist and the medical oncologist classifies the tumor as a low risk, HPV related tonsillar carcinoma cT1 cN0 vscN1 M0. No histological nor cytological examination seems feasible given the anatomical location of the node.*
Depending on nodal involvement, different approaches could be recommended.
Without evidence of node extension, trans oral surgery might be the best option, offering curative treatment without long term toxicity typically associated with radiotherapy. In the case of lymphatic spread, there is an advantage for radiotherapy due to its ability to dominate the para-pharyngeal

space.

Without clinical tool to support treatment decision, one might decide which treatment deliver sharing doubts and risks with the patient.

BD2Decide, collecting clinical and radiological information, could give the clinician a feedback regarding disease's risk of node spread, affecting treatment decision.

The BD2Decide models shall be run by the physician and the results displayed, The physician should have the possibility to zoom into the results and check patient's data before taking the treatment decision to be discussed in Tumor Board.

Determination of QoL at baseline

Before treatment start (-7- 0 days) QoL questionnaires will be distributed and recollected.

The questionnaires will be administered on paper/PC/tablet. The answers will be recorded. The overall QoL scoring will be also automatically calculated and recorded for comparison during post-treatment and follow-up.

A 67 year old patient with a carcinoma of the pharynx is referred to the hospital. After the clinical examination the physician submits to the patient a questionnaire (EORTC30 and EORTC35) to determine his QoL. The questionnaire is proposed on a tablet PC, where the patient can easily tick the answers to the questions. More complex questions are explained also with some clarification examples, to ensure that the patient correctly answers.

The results are recorded into the BD2Decide patient's data repository.

***User needs**. The QoL questionnaires shall be available in different formats: paper, laptop/PC and tablet. The questionnaires shall be provided in the language of the patient.*
For electronic versions of the questionnaire a contextual help should be available for all questions and especially for the more complex/unclear questions. The patient should be able to tick on the selected answers (score) and the score will be automatically recorded. The layout of the screenshot should be easily readable.
The questionnaire scoring should be available to the physicians for consultation.
QoL questionnaires should be collected periodically and all the results taken at different periods recorded.

Pathology examinations and data collection

Use case 1. The biopsies taken during the patient's first clinical examination are transferred to the pathology unit and examined by the pathologists to determine the presence of cancer and a preliminary staging (clinical staging, cTNM). If the histology is positive the biological specimens are stored as usual in FFPE for use. The biomolecular features indicated in the dataset (see Annex II, pathology data) are extracted and recorded (inserted by the pathologist into the patient's health records) for the patient.

The referrals of the pathology exams are added to the patient's health record. The FFPE specimens are then prepared as foreseen by the clinical protocol to be examined for HPV and genomic data extraction (qSeq).

Use case 2. A patient with Oral Cavity cancer undergoes surgery. During surgery the surgical tissues removed from the tumor lesions is taken to the pathology unit (within 30 minutes) and examined by pathologists. The margins are examined and the referral concerning margins invasion is immediately sent to the surgeons, so that they can eventually apply additional surgical procedures. The pathologist stores the tissue specimens for in-depth pathology and histology exam. After some days (usually 5-10 days) the referral is produces with the clinical staging, the grading and more histological and pathological markers (see Annex II pathology data). The referrals are added to the patient's Health Records. The biological tumor specimens are prepared and FFPE pieces are stored as for usual practice. The FFPE specimens are then prepared as foreseen by the clinical protocol to be examined for HPV and genomic data extraction (qSeq).

User needs. *In case the BD2Decide system is accessible from the operation theatre, the pathologist, after examining the margins, should be able to send an automatic alert to the surgeons on the BD2Decide system and inform the surgeons in case margins are affected/free. In all cases the information should be recorded in the BD2Decide dataset.*
After the result of the exams by the pathologist, the extracted data should, if a LIS is in place, be automatically transferred from the laboratory information system into the BD2Decide dataset. A user-friendly data entry should be provided to allow data transfer into the BD2Decide dataset.

If possible, the patient's ID should be automatically transferred either from the LIS or from the paper referral by means of bar code readers, in order to avoid errors.

The patient's data, after anonymization/encoding will then be stored into the BD2Decide anonymized data repository, using a code matching table stored into the hospital BD2Decide layer.

Pathology data should be approved (signed) by the pathologist.

Radiological examinations

Radiological staging

Use case. The patient with clinical diagnosis of stage III oral cancer is referred to the radiology department for pre-surgical staging. The referred patient will undergo CT and/or MRI and/or DWI-MRI for the assessment of radiological TNM stage and for the extraction of additional features. In particular, T parameter will be assessed on the basis of tumor size (e.g. cut off will be 2 and 4 cm in maximum diameter, as stated by current radiological staging system), diffusion to contiguous structures (e.g. cortical bone, inferior alveolar nerve, floor of the mouth, intrinsic or extrinsic muscles of the tongue, maxillary sinus, and skin), and more advanced diffusion toward masticator space, pterygoid plates, skull base, internal carotid artery. Otherwise, radiological signs of lymph node involvement will be used to provide N parameter magnitude. Notably, the short axis of lymph nodes will be used as parameter of lymphatic involvement, namely, lymph nodes larger than 10 mm will be deemed involved. Besides, the location of involved lymphnodes will be described according to the side and the anatomic level. In particular, ipsilateral or contralateral lymphatic involvement

will be defined on the basis of primary tumor side. Therefore, the anatomical level of involved lymph nodes will be reported according to the 7 level radiological system, including superior mediastinum.
Radiological signs of contiguous or lymphatic invasion will be used for the preoperative radiological staging. However, the accuracy of radiological signs is significantly lower than the reference standard, namely, pathological staging (pTNM). Thus, comparison between radiological findings and pTNM is ideal to refine radiologist experience and, hence, confidence in preoperative staging of oral cancer.

User needs. *The radiologists de-identifies the radiological exam for use in BD2Decide and applies the BD2Decide code. The radiologist will use Fraunhofer software to analyze the images of the tumor and the lymph-nodes and to determine the relevant parameters in a semi-automatic or in an automatic way. The parameters extracted complement the list of measurements done by the radiologist. Additionally the software performs tumor and lymph-nodes segmentations to determine volumes and axial measurements and to prepare the image for radiomics evaluation.*
Diagnostic images shall be anonimyzed prior to be stored in BD2Decide dataset (e.g. using the RSNA CTP anonymization software during upload. Images, if stored in BD2Decide dataset, should be suitably processed in order to prevent 3D reconstruction of the patient's identity.
The radiologist verifies the data extracted by Fraunhofer software and then, by means of a command button, transfers the data to the BD2Decide database. The processed images will also be stored for radiomics processing and possible legal issues.
The radiologist extracts a few significant images (jpg format) and send also them to the BD2Decide DB. These images are those visualized by physicians in the imaging data screens.

Image analysis software feature extraction

The image analysis software is designed to make the image feature extraction process of the radiologists as fast and convenient as possible. The software allows the radiologists to review the different image modalities of a patient. The image data can be analyzed, using the axial, sagittal and coronal views and a 3D reconstruction. The software is supports the creation of segmentations of the lymph nodes and tumor and guides the radiologist through the whole feature extraction process. Once all features are successfully extracted, the data can be send to the BD2Decide database for further analysis.

Use case 1: Patient with image data already on the clinical PACS. A patient that, who is already known to the hospital and for whom images have already been acquired, is to be analyzed. In this scenario, the patients' image data is already stored on the clinical PACS server. The software requests the image data of the specific patient from the clinical PACS server, which transfers the medical image data to the image analysis software where it can be processed. The extracted features can then be send to the BD2Decide database.

Use case 2: Patient coming from a different hospital. In cases where the patient coming from a different hospital, the image data is usually not present on the clinical PACS server. In this scenario, the user usually brings the image data on an external data storage with him. There will be 2

possibilities how to proceed in this case. The patients' image data can then be uploaded to the clinical PACS server, which would lead to Use case 1. Or the image data can be directly loaded into the image analysis software. Hereby, the image feature extraction process can be executed without the need of a clinical PACS server. Once the features have been successfully extracted, they can be send to the BD2Decide database or stored locally, to be send at a later point in time. This can be very useful if not active network connection is available.

User needs. The medical image data acquired from MRI or CT will be processed with the image analysis software. The radiologist will load the image data of the patient to be analyzed from the clinical PACS or an external data storage. The radiologist does the segmentation of the tumor and the lymph nodes with some automatic and semi-automatic segmentation tools provided by the image analysis software. Further, the radiologist determines if the tumor or the lymph nodes infiltrate any critical structures, as well as their primary location, side and some more features. The image analysis software automatically calculates the major and minor axis and the volume of the tumor and lymph nodes. As a last step, the software determines the t-staging of the tumor and the n-staging of the lymph nodes fully automatically. After the patient is saved the radiologist uploads the extracted image features to the BD2Decide database which can be achieved with a single click.

Radiomics features extraction (MAASTRO, INT, POLIMI)

The radiomics software applied to CT scans and/or MRI scans will allow to better characterize the tumor presentation in each specific patient, and will extract features that will be used to refine the prognosis prediction and allow clinicians to better understand the patient's prognosis and decide treatment.

Use case 1: Radiomics features extraction from CT/PET scans

After the PET/CT of the patient is acquired and segmented by the radiologist, using the software developed by Fraunhofer, radiological data are routinely collected. Patients' risk stratification and treatment decision also considers the complex heterogeneity of HNC tumors. To this end the radiologist uses the OncoRadomics software to analyse medical imaging.

The OncoRadiomics software allows the clinician to extract and analyze large amounts of advanced quantitative imaging features (e.g. tumor intensity, shape or texture) to quantify tumor heterogeneity. The results are then merged with the patient's clinical and/or molecular data in the BD2Decide Clinical Decision Support System (CDSS), based on which a report is generated with prognostic and/or predictive information regarding the patient and treatment options. The report is interpreted by the clinician (oncologist/surgeon), but also presented to the patient in an accessible way with the BD2Decide co-decision tool (e.g. with animations), which enables the patient to make a side-by-side comparison of all treatment options and pros and cons. The patient and the clinician then decide together which is the optimal treatment to select (e.g. the treatment that guarantees the best survival and quality of life for the patient).

User needs. *The segmented images processed by the radiologist either with or without the use of*

Fraunhofer software will be made available to the OncoRadiomics software. To this end appropriate image transfer/sharing (e.g. PACS) needs to be established. The OncoRadiomics software shall be called by the radiologist through the same interface (menu panel) used for BD2Decide radiological exams processing and imaging data extraction. The radiologist will use the OncoRadiomics software to extract the relevant radiomics features related to both healthy and cancerous anatomical areas. The data extracted will be transferred to the patient's electronic health records (see https://www.youtube.com/watch?v=Vf0F7q8vaS4*). The report (pdf format) generated by the OncoRadiomics software will also be stored in the BD2Decide data repository to allow consultation by any involved clinician and to be eventually presented to the patient.*
A specific data visualization should be provided to all clinicians for the radiomics images, the referral and the extracted features.

Use case 2: Radiomics features extraction from MRI scans

After the MRI of the patient is acquired and segmented by the radiologist, using the software developed by Fraunhofer, images are ready to be analysed by the Radomics software developed by POLIMI. Patients' risk stratification and treatment decision will be enriched by the info on complex heterogeneity of HNC tumors that can be extracted using POLIMI software.

The Radiomics software allows the clinician to extract and analyze large amounts of MRI quantitative imaging features (e.g. tumor intensity, shape or texture) to quantify tumor heterogeneity. The MRI radiomics data are used to feed the prognostic models and might be part of the personalized prognostic signature. MRI radiomics may be used when CT scans are not available or also in combination with CT radiomics data. The detailed radiomics features extracted by POLIMI software are then made available to the BIG Data Analysis to be merged with other data (e.g. genomics). The aggregated MRI radiomics features relevant for tumor characterization are are then merged with the patient's clinical and/or molecular data in the BD2Decide Clinical Decision Support System (CDSS), based on which a report is generated with prognostic and/or predictive information regarding the patient and may be used by the multidisciplinary team to take informed decisions regarding treatment options.

User needs. *The segmented images processed by the radiologist will be made available to the POLIMI Radiomics software. To this end appropriate image transfer/sharing (e.g. PACS) needs to be established. The POLIMI Radiomics software shall be invoked by the radiologist through the same interface (menu panel) used for BD2Decide radiological exams processing and imaging data extraction. The radiologist will use the POLIMI Radiomics software to extract the relevant radiomics features related to both healthy and cancerous anatomical areas. The data extracted will be made available to the BD2Decide data repository and to the BD2Decide Clinical Decision Support System (CDSS).*

Genomics and pathology biomarkers determination

While existing genomic signatures need to be further assessed, and BD2Decide specifically addresses this point, other biomolecular factors, such as HPV have already been assessed as

relevant for disease prognosis, in particular for OPSCC. Therefore we clinicians will need to determine all the relevant biomolecular markers, including HPV during the diagnosis phase. The following use cases apply for HPV+ and HPV- patients discrimination and for a personalized genomic profiling.

This functionality is automatically provided by the BD2Decide models and Big Data Analysis and it is recalled by the users anytime they invoke the prediction models or data analysis results from the BD2Decide platform.

Consult evidence medicine information

International guidelines such as ESMO-ESTRO-EHNS and all PUBMed publications relevant for Head and Neck Cancer should be checked by the Big Data system and relevant papers and information should be proposed on demand to the physician.

Use case 1. A 42 years old stage III hypopharynx cancer patient is eligible for surgery and post-surgery combined radio-chemotherapy treatment. However the patient is very young and does not have risk factors. The surgeon wants therefore verify in literature if some conservative treatments can be adopted for this patient which have proven effective. The physician therefore consults the PubMed and clinical guidelines by using the BD2Decide platform, from the patient's data screen. The BD2Decide system analyzed the available literature sources and the available patient's anonymized dataset and produces the following results:

- List of patients matching the same tumor localization, age, gender and staging of the patient, with the applied treatment and the outcome (if available) or the survival prediction. The physician can visualize all the available patients' data by clicking on each patient name.

- List of literature sources relevant for the case. The physician can consult the proposed papers.

User needs. *BD2Decide Big Data should analyze PubMed and clinical guidelines to search for relevant information for the specific patient case. Eventually the physician can propose a question for data analysis.*
The BD2Decide platform will produce a list of results as indicated in the use case and allow the consultation of data by the physician and the printout of results.

Access to public data for evidence-based knowledge

Use case 2. A clinical researcher is interested in correlating the BD2Decide data for larynx cancer with European data on cancer incidence from most recent surveys, in order to identify potential behavioural, age and gender related factors for the cancer incidence in Italy vs. Germany.

The BD2Decide Big Data analysis queries the epidemiology datasets at EU level from IARC (http://eco.iarc.fr/eureg/), OECD databases (https://data.oecd.org/health.htm) on health status, health risks (smoking, alcohol abuse), to the Embase database (http://store.elsevier.com/embase) and combines these aggregated data with BD2Decide data for larynx cancer cases. The risk factors are proposed with their relevance score.

User needs. *BD2Decide Big Data should analyze PubMed and available databases to provide aggregated information (probability or relevance scoring).*
The system should allow some basic querying to orient data analysis and the provision of results.

3.2 TREATMENT DECISION

Prognosis prediction

Use case 1. The surgeon needs to assess the risk of reoccurrence of the patient before treatment. After visiting the patient and after having assessed the need for surgery, the surgeon opens BD2Decide platform and visualizes the patient information on the main screen (control panel). Using the devoted button (or menu item) the surgeon activates the risk prediction model.
Different prediction models can be selected, which use only some of the data (e.g. only clinical data, etc.) or different combinations of data, or all available data.
The system shows on the screen the following information: main data regarding the patient including tumor localization, clinical and radiological staging, risk factors and clinical assessment of lymph-nodes. Upon request (button) the surgeon may visualize selected images. Then the model produces the following information:

- overall risk score for disease reoccurrence (and if possible timing of possible reoccurrence, depends if the model can produce this information)
- factors (bio-profile) which give evidence of the risk and specific value of each factor
- confidence levels
- graphical representation of the prognosis (bar chart) and Kaplan-Meyer curves, highlighting the patient vs. average population and vs high- medium- low-risk survival curves.

An estimate of a relatively high risk of bad prognosis by additional parameters such as genomics data and imaging data, together with the finding of the focal non-cohesive growth pattern of the tumor, might have directed the decision to adjuvant post-operative radiotherapy. An estimate of a low risk for recurrence by imaging and genomics parameters, the age of the patient and the aspiration problems in the past may have led to the decision not to threat this patient by adjuvant postoperative radiotherapy.

When any information of the patient is collected to decide about the patient´s therapy, the pre-therapeutic tumor board takes place. While treatment, e.g. after surgery and receiving histopathological data of the resected tumor and lymph nodes, the prediction could help to modify and optimize the further treatment options. The clinician therefore must see how the new data influences the prior prediction and which central element accounts most for a possible change.

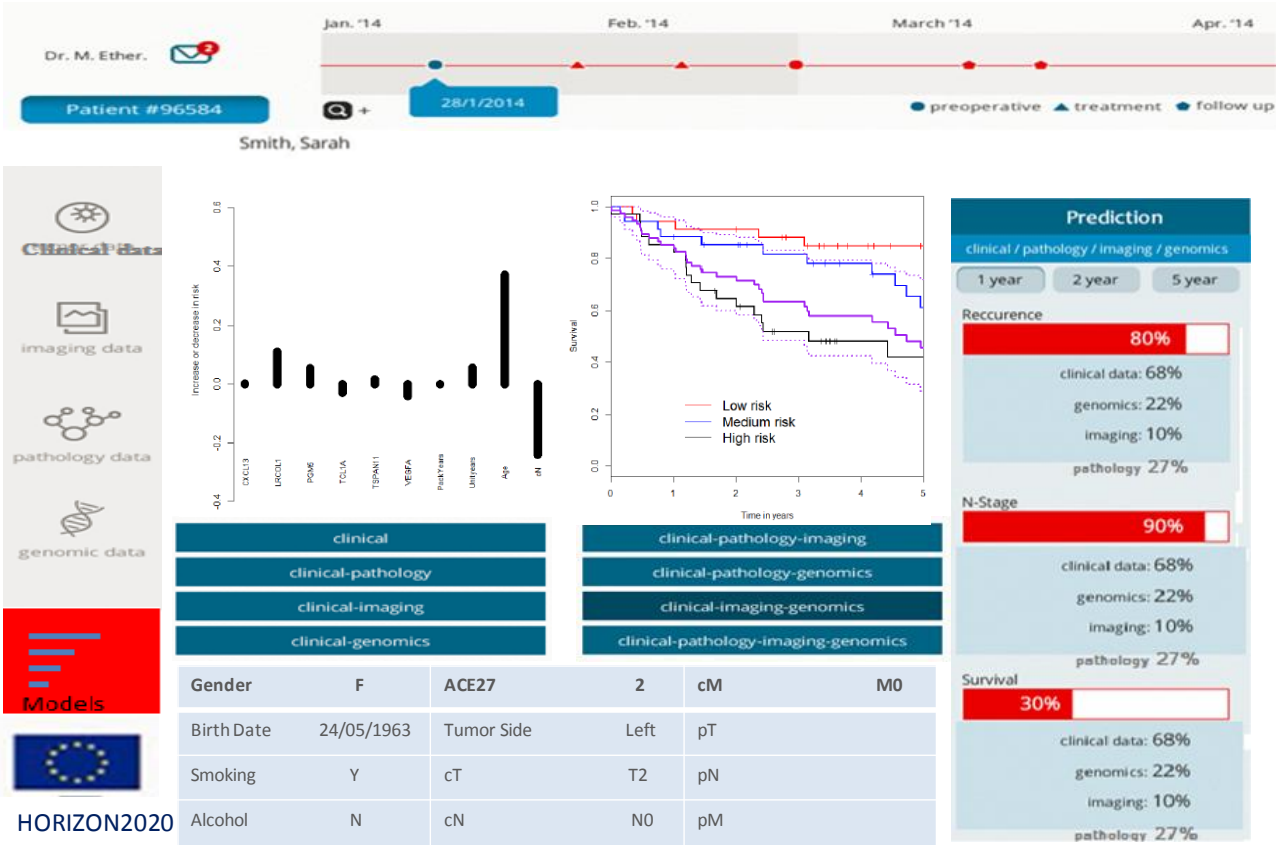


Figure 6. Example of visualization of patient-specific prediction (courtesy OraMod project)

Note: New data is any new data coming up while the patient receives diagnostics and treatment. Here it might be the final histology after surgery that may contain information about extra-nodal spread, or close margin resection that may influence the prediction. The MD should see what kind of additional information compared to the data set existing before, has influenced the prediction (in this case the lymph node status and the close margin resection).

User needs. *The surgeon (clinical oncologist, radiation oncologist) will run the risk prediction model. The system will provide the prognosis (survival) estimation based on the selected model. The surgeon must be allowed to see the different factors which concurred to the risk estimation and have different views of the patient's data. If it is the case, the surgeon may perform simulations (i.e. run different models). To facilitate comparisons, the different prediction results shall be opened in different page layers (or the data such as % and contributing factors should be visualized in the same page, for immediate visual comparison), while graphs might be requested for each specific prognostic model.*

If different kinds of information are taken into account for calculation, the relative % of their prediction value must be visible (e.g. in a pie chart).

Comparing similar cases

Use case. The clinician has a very particular case to be studied. He accesses the BD2Decide system and selects the patient to be studied and indicates the specific characteristics (i.e. T-Localization, TNM staging, age range, sex) and asks the system to perform a verification on similar cases. The system retrieves cases with similar parameters and shows the data regarding outcome at date (last follow-up), the applied treatment and the tumor characteristics, and the reoccurrence risk for each case. A graph with the different reoccurrence risk estimation for the study patient compared to the other similar cases is proposed. By clicking on the graph, the details of each case are shown.

The clinician wants to simulate the patient's outcome based on previous cases. He selects the data to be analyzed (age, risk factors, site and size of the tumor, imaging, histology, surgery, chemoradiation, etc.) and the system re-calculates the outcome prediction (probability of reoccurrence, no reoccurrence) in form of graph or map, for the case to be studied and for the similar cases extracted from the BD2Decide data base.

The clinician should also be able to select which parameters to evaluate, in order to assess how a single parameter influences the patient's outcome (example: 40 years old patient, with T3N2aM0 staging. I would like to know outcome of other patients in the same condition and understand which are the most critical factors influencing the prognosis for example adding smoke in the parameters evaluated). Finally in case patient's outcome differs from other similar cases, the system should be able to analyze which parameters lead to this difference.

***User needs**. Regarding a group of analyzed patients it should be possible to correlate single data sets with the reoccurrence risk on the one hand but also with other data sets on the other hand. By the possibility to correlate the data, subgroups of patients with special risks might be identified. Furthermore, data sets with a high or low correlation (e.g. clinical data and genetic data) within the system can be depicted. These functions should also be illustrated by graphs, e.g. correlation maps.*

Patient's stratification and clustering by risk

Use Case1. A physician of the multidisciplinary team has a particular oral cavity cancer patient for whom a treatment decision must be taken. The BD2Decide DSS and the models and Big Data analysis have provided the prognostic predictions but the clinician would like to better characterize the individual patient inside a cluster of similar cases. The clinician enters BD2Decide data analysis functionality and selects the set of prognostic factors (TNM, age, gender, race, tumor localization etc.) and risk factors (smoke, alcohol, nutrition, oral hygiene, infections). The system presents:

1. the graphical representation of the patient's prognosis as compared with the prognosis of a similar subgroup (cluster) of patients with the same characteristics (prognostic factors, risk factors)

2. highlight similarities between the individual patient and the other cases in the "cluster".

This functionality will be used to support clinical decisions and apply personalized medicine.

***User needs**. The system shall provide a functionality that allows the physician and/or the researcher to select specific factors relevant for one target patient and extract prognostic*

information of the patient as compared to a subgroup of similar cases. The system should use the DSS functionalities (e.g. models) to produce the prognostic prediction in both quantitative data representation and in graphical representation (e.g. survival curves) and also produce a list (and i possible also a plot) of the similarities between the individual patient and all the identified similar cases.

Risk stratification and survival prediction simulations

Use Case: The clinician is not convinced of BD2Decide prediction results and decides to make some simulations. Therefore the clinician selects different models and asks that they are run and results shown in a comparative way. The TNM prognostic system should always be shown as reference.

It should be possible to do a prediction only regarding single or few combined data sets, like e.g. comparing the reoccurrence risk regarding only genetic data compared to regarding only clinical or only imaging data or a combination of the last two. Comparisons should appear on one screen.

User needs.
The clinician must be able to see if his decisions are taken according to BD2Decide or standard international guidelines (TNM) because he must be aware of what leads to his final clinical decisions concerning the patient and his life. The gold standards may be included to be retrieved by the clinician upon request (pdf documents).
Note: These guidelines are also constructed like a kind of "decision scheme". In certain clinical situations, certain treatment decisions are mostly recommended. It should be possible to include fixed decision trees based on guidelines. The necessary information for the decisions based on common guidelines and therefore accounting for decision trees is anyway collected by BD2Decide.
The reoccurrence risk is calculated by the time of the first histo-pathological diagnosis based on a normally small tumor specimen. Only limited histo-pathological characterization is possible at this point and limited clinical and imaging data is available. Genomic data also follows later. Patient´s demographical and clinical data can be used for prediction and the clinician has the possibility to see how the risk prediction changes with the input of information. The clinician follows the risk prediction and can see how any additional information gives changes to it. So e.g. when imaging is performed, he can see how the additional information changes the reoccurrence prediction.
All BD2Decide predictions should anytime be compared to prediction regarding common risk factor that must be defined as one fixed data set that always appears in one corner of the screen if prediction is asked. Data belonging to this "traditional data set" is e.g.: patient´s age, sex, risk factor (nicotine, alcohol), TNM, HPV, extracapsular spread, vessel/bone/scull base infiltration.
Regarding one specific patient a timeline should be visualized that shows, how prediction changed within the course of illness when more and more data was added. Therefore it is shown how information but also treatment decisions might influence the prediction.

Patient's case discussion (Tumor Board) and treatment decision

The clinician needs to discuss a specific case in a Tumor Board (i.e. with the colleagues of other medical disciplines). This may be done during the weekly Tumor Board (physical or virtual meeting) or by means of an exceptional Tumor Board virtual meeting requested by the surgeon. The tumor board consists of head and neck surgeons, oral and maxillofacial surgeons, head and neck radiologist, radiation oncologists, medical oncologists, pathologist, oncology nurses (if applicable), and residents in training. Together they discuss the patient initially, and after panendoscopy and collection of all histological and radiological data, to decide on the treatment plan. This would be also the moment to bring in the genomics or radiologic data.

Use case 1: **Tumor Boards planning.** The Tumor Board chair defines the dates of the next Tumor Board Meetings and records them in the system. Once the Tumor Board is planned, both the Tumor Board chair or the individual clinicians can enlist patients to be discussed.

User needs. *BD2Decide must foresee a functionality that allows planning of Tumor Boards. An agenda opens and the Tumor Board date, time, location and participants can be indicated by the Tumor Board Chair.*
Once defined, all participants will receive an automatic invitation which could be added to their online agenda (e.g. Outlook agenda).
Participants will confirm their participation by checking the relevant participation checkbox. Each participant may include patients to be discussed in the patients list of each Tumor Board.
A reminder of the Tumor Board meeting can also be indicated by each participant, to be issued before the date/time of the meeting.

Use case 2: Tumor Board invitation and patients discussion. The surgeon enters BD2Decide platform and opens a Tumor Board request, inviting all colleagues. In the request the clinician indicates the case(s) to be discussed and opens a session for each case to be discussed, where on a blank form the case is presented. The Tumor Board meeting date and time is proposed or an asynchronous discussion is agreed (i.e. no simultaneous videoconference, but each clinician inserts into the tumor Board form his/her considerations regarding the case to be discussed). If a videoconference is planned, a message is sent via sms and email to participants to invite them. At the tumor board time, the surgeon opens the videoconference from BD2Decide collaborative decision making platform and all Tumor Board members can access the patient's data and discuss the case. The minutes of the discussion are recorded. Before the Tumor Board all involved physicians can visualize all patient's data, imaging and run the prognostic model from the BD2Decide platform.

The data and decisions are electronically reported in a so called CONSOLIDATED REPORT form. Letters with the relevant findings are sent to the referring physician and the general practitioner. Additional histo-pathological and genomic data will be reported by the pathologist and additional radiologic data by the radiologist. The resident in training presenting the patient will evaluate these data and bring in the BD2Decide risk assessment and prognostication.

User needs. *BD2Decide will offer the possibility to support either a physical or a virtual tumor board. In the second case the participants may participate:*

- *synchronously (i.e. at the same time of the physical meeting, but in a different location) through a video or voice conference provided by BD2Decide, or*
- *asynchronously, i.e. in a previous day, and in this case they must record their opinion in the Virtual Tumor Board session screen*
- *physically during the meeting. in this case their opinion is recorded by the person chairing the case discussion.*

The software must allow to invite all Tumor Board Participants and to select all the cases to be discussed. Invited clinicians will receive an email and a SMS on their mobile phones with the date, time, location and list of patients to be discussed. The list of patients to be discussed will be created either by the Tumor Board chair, who selects the patients from the BD2Decide dataset, or will be inserted one by one by the clinicians in charge, after the patient's visit and first diagnosis. In this case the system should allow registration of the patient for the next tumor board discussion (e.g. the clinician asks for this inclusion through a button on the patient EHR. The system will automatically enlist the patients in the Tumor Board as soon as the date and time of the next Tumor Board are defined.

Data from the patient must be visible to all Tumor Board members on a shared virtual space (conference call plus shared blackboard for data visualization especially imaging). The statements of each participant will be registered, by inserting them directly in the Virtual Tumor Board screen.

The results of the tumor board (treatment decisions) must be recorded (and eventually printed, upon request). The Tumor Board members must digitally sign the tumor board decisions.

Patient's-Physician co-decision

The clinician communicates to the patient the treatment decision taken during the Tumor Board. The consequences of the treatment and the possible prognosis are discussed. At present the discussion is made based on the clinicians' experience. With BD2Decide the clinician has two main tools supporting the communication to the patient:

1. the prognostic models, that provide data and probabilities concerning the patient survival

2. the co-decision aid tool, that presents to the patient the specific tumor characteristics, the possible treatment options with the pros and cons and allows the patient defining his/her preferences especially for what the different treatments can impact on personal engagement and commitment and quality of life. The goal of the Decision Aids Tools is to assist cancer patients in making an informed decision about which treatment would suit them better.

Use case 1: The patient is Mrs. Holland, who has been diagnosed a Stage III Oropharynx cancer. The clinician in charge is Dr. Smith, an oncologist. After Tumor Board the most appropriate treatment proposed for Mrs. Holland is radio-chemotherapy. This might extend survival for 2 years, according to survival predictions. Dr. Smith receives Mrs. Holland to present the treatment plan. The treatment options for her tumor type and the selected treatment pros and cons are presented on

a tablet, by means of videos and animations. Mrs. Holland can individually look at the treatment options and understand their characteristics, side-effects and implications in the short and long term. She can also compare the treatments side by side. This will help her to understand why, compared with other treatments, chemo-radiotherapy would improve her quality of life and extend her survival. Mrs. Holland can directly indicate on the tables screens her preferences/opinions when she is presented impacts of the treatments on number and frequency of treatment sessions and impacts on quality of life or limitations to usual life habits.

Mrs. Holland can ask Dr. Smith clarifications, and get a digital or printed copy of the treatment characteristics and side effects for future reference. After reviewing the information she finally agrees with Dr. Smith on the proposed treatment.

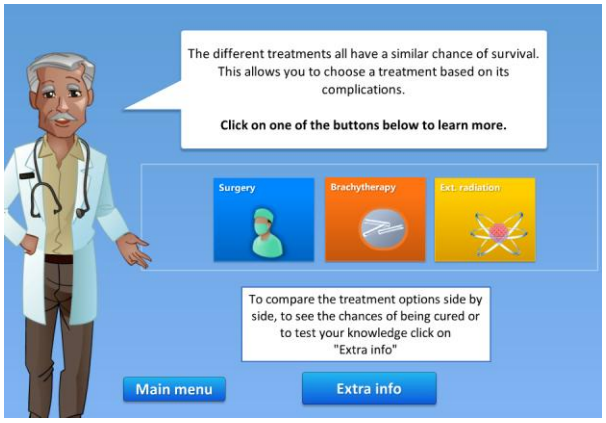
User needs. *BD2Decide must foresee the activation of the on-line co-decision aid tool. The tool can be used by the patient alone or jointly with the clinician. In this case the patient operates the tool and the clinician only answers to the patient's questions.*

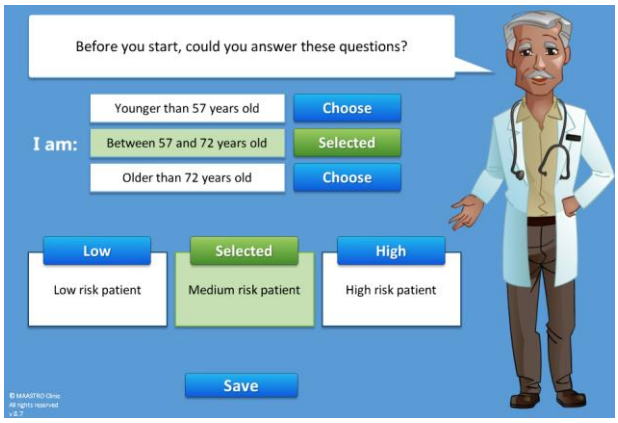
The tool will automatically get the patient's data from the BD2Decide patients records dataset as required (e.g. age, tumor localization, stage etc.) and the risk level as calculated by the prognostic model and show the data for confirmation. The possible treatment options will be shown and the proposed option after Tumor Board decision is selected. This option may be changed by the patient so that he/she can compare the different options.

The tool will present some questions. By answering these questions, patients can see the pros and cons of the treatment and decide which treatment suits them better.

Once started, the software tool offers several options (see screenshot below courtesy of MAASTRO, from existing aid for prostate cancer. Animation: www.treatmentchoice.info go to prostate cancer English password: Maastro123 or go to lung cancer English).

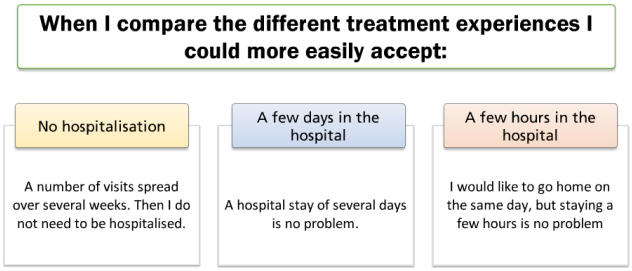
The patient can see general information concerning the specific disease and/or proceed to the decision support aid. For each treatment a description (supported by images/videos), list of pros and cons, list of recommendations, side effects and complications can be visualized.

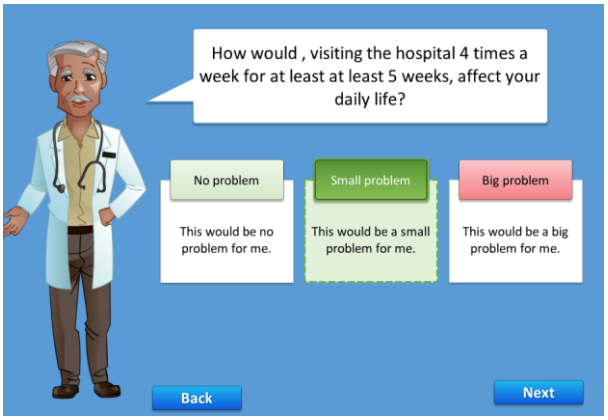




The survival/disease free survival chances for the different treatments are also proposed (based on literature). NB. The survival options for the patient as calculated by the models are ***not shown*** *(they may be illustrated by the clinician upon request).*

The patient can select his/her preferences (length of treatment, acceptance of side effects etc.)





and the results in terms of survival and complications are shown and can be discussed with the clinician and can be printed.

The tool should be available also over the web, independently from BD2Decide. In this case the patient will need to insert some data.

Planning of treatment

Use Case 1. The clinician wants to simulate some therapeutic scenarios and verify how they would influence the risk prediction. To this end a comparison with previous cases may be done, with same tumor localization and staging and different therapeutic approaches.
Therefore, the tumor board participants could see how the potential outcome could be, if different therapeutic scenarios would take place, e.g.:

- only surgery (possibly with different outcome scenarios)
- surgery and additional radio(chemo) therapy
- radio(chemo)therapy alone
- different kinds of radio(chemo)therapy with/without surgery

Note: If BD2Decide allows comparing the current case with similar cases already included in the system and the related outcomes, then the system itself would provide the scenarios and the outcomes.

***User needs**. BD2Decide should allow users to activate this functionality. The system should select cases matching the current patient by tumor localization and TNM stage (clinical staging or pathology staging if this is available). Then the system should run the prediction on such cases and propose the risk prediction calculation, depending on the applied therapy (in graphical format), compared to the reoccurrence risk of the patient under study. The clinician should be able to compare all scenarios on one screen with a nice graphic, illustrating the patient's survival prediction compared to other patient´s cases (e.g. dot cloud or field with graduated colors according to expected risk or both together in one image).*
It would also be nice to have some suggestions of the system showing patients with comparable parameters, who have been treated before, together with their actual outcome. Then known old cases could be compared and discussed with the new case and the expertise generated by the knowledge of the "old case" could be taken into account. The clinician therefore indicates that

such a simulation is requested and the system proposes a list of matching cases, with different therapeutic options, and the related reoccurrence risk estimation.
Furthermore, such a tool would support a teaching effect. Possibly data predominantly affecting the survival risk and also data possibly negatively influencing the accurate prediction could become conspicuous with this function, thus be identified and emphasized if relevant or indicated as non relevant within the system.
At any time point, the consequences of data integration on the prediction of survival must be obvious. The BD2Decide Prognostic prediction must be set into context with common guidelines and both must be visible for treatment and follow-up decisions for the clinician at first sight.

3.3 TREATMENT

Use case 1. The choice made after the Tumor Board is for a composite resection with selective neck dissection. The defect is reconstructed by a free flap of the forearm. The excised specimen is sent to pathology. The findings are a T3 squamous cell carcinoma with a single lymph node metastases: pT3N1. There is no extranodal spread of the neck node. The tumor is completely resected with a margin of 8 mm. However, both vaso-invasive and perineural growth are reported, and there is focal non-cohesive growth pattern of the tumor. In this case the difficult decision is on adjuvant post-operative radiotherapy. The patient is 65 years old, in general good health, but had some swallowing aspiration problems in the past.

User needs: All the surgery procedures performed must be recorded and a surgery report printed. Data from surgery which are mandatory (for legal reasons) must be collected with the same interface (link to existing software) and sent to the HIS (for administrative purposes).

Also the report coming from the diagnostic imaging and the pathological report should be standardized and in agreement with guidelines. In the study we performed we noticed a very heterogeneous reporting often inadequate and/or incomplete

Patient's case study by means of full data visualization and simulations and different perspectives

Use case. After surgery, the clinicians evaluate the case correlating together the data of the patient. Tumor board members need to assess simultaneously different parameters to give indication for further treatments. For example we need to know if that lymph node visible on the MRI has been resected and if there is extra-capsular spreading on the pathologic report, or we want to visualize on MRI or CT if there is any margin positive at the pathologic examination and where are these. To prepare Tumor Board the clinician accesses the BD2Decide system and asks for the cases to be discussed in Tumor Board. The system produces a list of cases with some data (i.e. T-Localization, staging, type of surgery) and the clinician selects the first case. The system produces an overview of patient's data (age, sex), of tumor and lymph-nodes information (Staging, localization, status of margins) and of the applied surgical procedures and visualizes the images selected by radiologists from CT/MRI scans. By clicking on the images, the clinician can view the relevant data of tumor, lymph-nodes and other anatomical parts. The clinician can also visualize the details of treatment, by zooming into the treatment or he can simulate the reoccurrence risk prediction (see previous use

cases). Finally the clinician can open a Virtual Tumor Board session and insert his comments regarding the specific case to be presented to the colleagues in the next tumor board.

User needs. *It should be possible to reevaluate the information details of any kind of data quality. So e.g. if the clinician clicks on the data block "imaging", the software should be able to show 1. the images themselves, and 2. the data generated from the imaging (tumor or lymph node size, localization, etc….). So the clinician can easily see the details and at the same time or with one click in between the proportion of value of this kind of data for the reoccurrence prediction.*

3.4 FOLLOW-UP

Follow-up visit

Use case 1. The patient arrives for follow-up visit. The clinician inspects the patient and eventually looks at CT/MRI scans available on BD2Decide platform Data from follow-up visit are collected and next visit is planned. Eventually the survival probability is re-assessed using BD2Decide prognostic models.

User needs. *After therapy, the system should propose a schedule for follow up care and generate a plan that can be printed out for the clinician as well as the patient. This plan should integrate clinical investigations as well as laboratory and imaging investigations. The plan must be changeable if any additional data modifies the prognosis. It could, e.g. be possible that a lymph node metastasis appears 8 weeks despite a low reoccurrence prediction by the system. Then the software must "react" and intensify the follow up and e.g. suggest a new imaging.*

Note: The rules are normally the fixed schemes for patient´s aftercare (follow up visits every 3 month, regular imaging etc.). Some clinics also have little booklets with fixed aftercare programs to be handed out to the patients.

Use Case 2. The clinician performs prognostic evaluation reassessment during follow-up
During the course of illness it would be nice if the individual prognosis is changed in a timeline, also visualizing the data input. The clinician runs again the prognostic models and verifies if the prediction is modified.

User needs. *The system shall allow re-assessment of prognostic prediction anytime during follow-up. Time dependant variables may be included (such as time after remission), to check if they affect reoccurrence risk.*

Determination of QoL during follow-up

Use case. During follow-up visits taken at month 6, 18, 24. QoL questionnaires will be distributed and recollected. The questionnaires will be administered on paper/PC/tablet. The answers will be recorded. The overall QoL scoring will be also automatically calculated and recorded for comparison during post-treatment and follow-up.

A patient treated for a carcinoma of the pharynx presents for a follow-up visit. After the clinical examination the physician submits to the patient a questionnaire (EORTC30 and EORTC35) to determine his QoL at the time of follow-up. The questionnaire is proposed on a tablet PC, where the

patient can easily tick the answers to the questions. More complex questions are explained also with some clarification examples, to ensure that the patient correctly answers. The results are recorded into the BD2Decide patient's data repository.

The clinician can revise the actual questionnaire overall score and compare it to the scores of previously collected questionnaires.

User needs. *The QoL questionnaires shall be available in different formats: paper, laptop/PC and tablet. The questionnaires shall be provided in the language of the patient.*
For electronic versions of the questionnaire a contextual help should be available for all questions and especially for the more complex/unclear questions. The patient should be able to tick on the selected answers (score) and the score will be automatically recorded. The layout of the screenshot should be easily readable.
The questionnaire scoring should be available to the physicians for consultation and comparison with previously collected QoL questionnaires for the patient (baseline, post-treatment).
QoL questionnaires should be collected periodically and all the results taken at different periods recorded.
A functionality to summarize and report (graphically and in tabular format) the QoL modification overtime for the patient and for subgroups of patients (e.g. same staging and T-Localization/tumor subsite) should be provided.

Patient's data visualization during follow-up visit.

Use case 1. The clinician must visit a patient during follow-up. It is the annual follow-up visit after the first year from remission (the surgeon sees this in the list of patients to be visited, from the BD2Decide control panel/daily visits list) and the surgeon needs to review the data of the patient before seeing him. Therefore the clinician opens the patient's EHR from the list of patients to be visited (or from the control panel).

The visualization screen shows the basic information of the patient:

- socio-demographic
- risk factors
- tumor and lymph-nodes data
- treatment data
- a subset of diagnostic images if available from radiology
- Follow-up visits and status
- survival prediction estimation (calculated at baseline).

The clinician can then visualize details of the patient record by zooming in into specific datasets (for example to verify the histology data, in order to assess additional risk factors or potential aspects to be checked during the visit).

Use case 2. At any moment the clinicians must be able to visualize the data of the patient. For example at each visit the surgeon must have the possibility to select the patient from the list of the patients to be visited this day and to view the patient's data based on the relevance of the data, with the possibility to zoom-in (i.e. go in more details) if necessary to better understand the patient's status. The clinician has a patient to be visited and so he opens BD2Decide and searches for the patient. He opens the patient's folder and sees the basic data:, and a timeline indicating the workflow events performed by the patient (i.e. baseline, 2nd visit, treatment, follow-ups ...). The clinician then can scroll through the different events/visits and visualize the patient's data. The clinician wants to verify the date related to lymph-nodes. The clinician may also use the visual tool, where the head and neck of the patient is shown (i.e. one image from Image analysis tools, selected by the radiologist). The clinician clicks on one of the lymph-nodes and the system shows the data related to this lymph-node and of the other lymph-nodes (if relevant). By clicking on the tumor on the image, the system shows all data related to the tumor (Staging, localization, genomic, etc., including imaging parameters, histology). Using a 'zoom' option the clinician can look at the details of each subset of data (such as imaging, histology, etc.).

Use case3. Patient's overview. A useful functionality is to show to each clinician the patients control panel, where all patients that are under his/her management are shown.

The control panel may be proposed in different situations:

1. view all HNC patients managed in the hospital

2. view only current patients (i.e. still in treatment/follow-up)

3. view only patients managed by the clinician who logged in

4. view all patients to be visited that day

For each visit the available data are listed, so that the clinician can have an overview of the patient status. We suggest that in this overview the reoccurrence should be visualized.

It is important that data are organized in the same way as they are collected (in case of data entry) and in a different way (i.e. aggregated by interest/specialty or by point of view) when proposed in visualization.

The overview might also contain a reminder function (highlighting) for patients for whom the follow-up visit is delayed/has not been implemented within the correct timeline (e.g. 3 month). Viewing this reminding function, the clinician is able to react and to e.g. contact the patient to initiate the follow-up visit. An additional function would be to create a list of all patients with delayed follow-up visits/ delayed imaging, etc.

User needs. *Since clinicians are neither statisticians nor IT specialists, the software must edit and illustrate the data in a simple, clearly arranged setting with a user-friendly interface and functions, limited to few necessary buttons. For this purpose, visualization of data using comprehensible images and graphics is mandatory. The first dimension that should appear when opening the software tool, is an overview about the patients main data set concerning him and his tumor (ID, age, gender, TNM, HPV, alcohol, nicotine (pack years), Grading, extranodal spread (Y/N), perineural infiltration(Y/N), vessel infiltration (Y/N), bone infiltration (Y/N)).*

Upon request, there should be a risk assessment visualized in form of a graph where the patient´s risk is pictures in relation to the other patient´s risk (i.e. patients with same T-localization and/or TNM staging). For comparison the system should visualize all patients integrated in the system where the reoccurrence has already been predicted and positively verified by the outcome. Furthermore it should be possible to see only patients in a graph who are identical in a subgroup of data sets and compare them in respect of other data e.g. reoccurrence risk. So the clinician can e.g. select all T3-staged patients and see their risk stratification or visualize which proportion of them is e.g. HPV negative or has special genetic expression profiles.

The kind of data that leads to the prediction (demographic, main tumor characteristics, imaging, genetic) and its individual proportion and value for the prediction is illustrated e.g. in a pie chart. See example from PredictAD project (www.predictad.eu/) illustrated here: http://files.kotisivukone.com/predictad.auttaa.fi/tiedostot/newsletter/predictad_newsletter2_2011.pdf).

Both graphs can be interactive and by clicking on patients with a similar prediction (e.g. other dots in a dot cloud) value like the investigated patient, the data of different individuals can be compared. Be clicking on the specific data subsets in the pie chart, that illustrates the percent proportion of this specific data set, the data leading to the prediction becomes visible.

3.5 REPORTING

The physician involved in patient's management will access the BD2Decide system to get some useful reports.

Use case 1. The clinician asks for a printout of the electronic patient's data. This function is activated by clicking on the "Print EHR" option present at the level of each patient. The clinician may select the sets of data to be printed (i.e. only patient's visits without referrals, all the EHR including referrals,...). The system prints the patient's EHR and includes in the printout all data available, grouped by time slice (i.e., visits, hospitalization, surgery, follow-up, etc.) and, if requested, in separate sheets the referrals from radiology, histology, the minutes of the tumor board etc.

Use case 2. The clinician is discharging a patient. He enters the system and authorizes discharge through a specific functionality. The system shows a form where the clinician enters the letter of discharge and signs it. The letter is then saved to be printed when the patient is dismissed from hospital later that day.

Use case 3. Consultancy report. A consultancy is performed on a specific critical case and the consultant clinician needs to produce an opinion report. The clinician enters BD2Decide for the patient, opens the "Consultancy" functionality and a form is presented where the opinion on the studied case is written. The opinion is saved in the system along with the name of the clinician and the date and time of editing. The opinion paper is printed and signed.

Use case 4. Additional reports. In particular the user interface will provide the following list of reports:

- Patients list by TNM staging: since today the TNM staging system is the most reliable parameter for prognostic estimates in patients affected by OSCC, this report will provide a list of patients with/without recurrence classified by TNM staging. The clinician will have the possibility to select the hospital and the level of detail of the document. If a "specific hospital" will be selected, only patients from the specific hospital will be extracted. Otherwise, if the choice "all patients" will be indicated, all patients in the local BD2Decide DB will be analyzed. Finally if only "totals" will be selected, the report will not present the detailed list of patients. The report will sort patients by TNM staging in this order: pT , pN and pM. Totals will be provided at each break of pT and pN. Totals will be provided either for each single hospital or for all hospitals. Data should be also available for export (csv, xls, xml formats).

- Patients list by risk factors: this report will present to the physician the list of all the patients with a specific risk factor and will identify which (and how many) had a reoccurrence, a persistence of disease or a complete remission at the date of the report. Also in this case the user interface will allow the clinician to select the hospital and the level of detail of the report. The clinician may draw on the screen the drop-down list so as to choose the specific risk factor to investigate. The totals will be calculated for disease-free patients, dead patients, and relapsers. Data should be also available for export (csv, xls, xml formats).

- Patients list by age group/age range: the software will provide the possibility to stratify the patients by age range and will calculate in each sub-group the total of relapsers and non-relapsers. The report will be done for each single hospital or for all the patients present in the local DB. Data should be also available for export (csv, xls, xml formats).

- Patients list by T-localization: this report is similar to the TNM staging one, but in this case the patients will be ordered by T-Localization: apparently redundant, this stratification is justified by the recently proposed theory that genomic aspects are strongly related to the localization of the tumour. The report will be provided either for a single hospital or for all hospitals. Data should be also available for export (csv, xls, xml formats).

- Patients list by surgical procedure: the surgeon will check the treatment options adopted for patients affected by HNC classified either by TNM staging and only by tumour site. The user will decide if the report should include only patients from a specific hospital or all the patients in the DB. Data should be also available for export (csv, xls, xml formats).

User needs. *The system must foresee the printout of all official documents (such as discharge letter, report on applied surgical procedures, patient's EHR etc.). Additionally the system should allow editing and printing of external consultancy opinions and of lists of cases to be analyzed by the clinicians. The system should be flexible enough to allow the clinician select the patient's extraction/grouping criteria and to indicate the data to be presented in the reports.*

Reports should be available on screen (reduced format with the most important data to be visualized on the screen) and on paper.

It could be useful also to be able to make simple statistics of patients treated such as treatment suggested by BD2 decide and treatment agreed with the patients, treatments and survival, treatment

and QoL. The use of BD2 can because a prospective cohort to use to reassess the prediction and adjust it in future.

3.6 INCLUSION OF POPULATION DATA AND OF EVIDENCE MEDICINE INFORMATION

3.6.1 Types of data to be included

Note: The INT and ISS can provide the data described in the table because these data have been previously collected for other studies. Concerning Cancer Registries (CR) data, ad hoc agreement should be stipulated with the CR in each country or with the European network of Cancer Registries (ENCR), in order to obtain additional data in future (i.e. after the end of the project).

The same applies to data provided by National Health Authorities through ISS (e.g. medication data, epidemiology data, lifestyle behaviours at population level) for data updates after the end of the project.

The population data are described in Annex III.

Data should be high-resolution as much as possible. Aggregated population data (e.g. by gegraphic area) might also be used when high-resolution data are not available (e.g. in case of environmental or population lifestyle data).

The life expectancy of the populations considered may be used in consideration of missing data on comorbidity.

Tumor registry data

Population-based cancer registry can provide data on patient (data of birth, country of residence and gender), tumor characteristics (histology and topography), Vital status. Some cancer registries have information also on hospital of diagnosis, on main treatment and hospital of treatment and on extend of disease. Cancer registry can provide a real picture of the prognosis of head and neck patients including those treated in expert centers and those treated in other kind of hospitals or not treated at all. More important cancer registries can provide the vital status. For registry with a long history, for patients is available the presence of other cancer, before and after the H&N tumours. Information on stage is not always available in the clinical record. For head and neck patients, to reconstruct the stage on the basis of the diagnostic imaging and of the biopsy might be complicated. Thus, it could be very important for cancer registries to access to this data via the BD2Decide. Also the definition of the sub-site might be challenging. Thus a confirmation on the site available in BD2Decide can be of interest for cancer registries

To avoid any issue related to the privacy, it could be enough o include in the patient consent form that the information will be available also to population-based cancer registries for epidemiological surveillance.

All data described in Annex III are electronically available and can be extracted in any convenient format.

Demographic and Lifestyle behaviours data

Demographic indicators such as: mortality and fertility rate, life expectancy, dependency rate (population 0-14 and 65+ divided by population 15-64), elderly rate are available by calendar year at aggregated level (census area or city of residence) at population level from the Italian National Institute of Statistics (http://demo.istat.it/).

Information on lifestyle risk factors and preventive measures such as smoking habits, alcohol consumption, dietary habits, overweight, physical activity, vaccinations, preventive health screenings are available since 2006 at aggregated level (health local unit) from surveys administered to representative samples of the Italian population (for example, the PASSI study: http://www.epicentro.iss.it/passi/).

Health data of the population and access to medication

Hospital Discharge Cards (HDC) database: includes information on all discharges of Italian residents from hospitals of the Italian National Health System. The information is collected at individual level and includes: demographic variables (age, sex, city of residence …); clinical variables (main and secondary diagnoses, main and secondary treatments); administrative variables (type of admission: ordinary or day hospital, date of admission and discharge, diagnostic related group, length of stay, cost of the treatment …). The HCD database is available in electronic from 2001 to 2014, the HDC are provided from the hospitals to the corresponding regions, the regions transmit the information to the Ministry of Health (http://www.salute.gov.it/portale/temi/p2_4.jsp?area=ricoveriOspedalieri).

Outpatient records: include information on all treatments and procedures administered from local ambulatories of the Italian Public Health System. The information is collected at individual level and includes: demographic variables (age, sex, city of residence …); clinical variables (treatment or procedure administered to the patient, related cost). The outpatient records are available in electronic format since a different period, depending on the Italian region, and are provided from the local health authorities to the corresponding regions.

Drug prescriptions: include information on all drugs prescribed to the patients from general practitioners and dispensed to outpatients from local community pharmacies or, in some cases, directly from local health units. This information is recorded at individual basis, and includes demographic variables (age, sex, city of residence …) and drug information (marketing authorization code indicating the specific formulation, date of the prescription, number of packages and price). At regional level, data on drugs dispensed to patients during the hospitalization are available. This information is aggregated at drug level and includes: marketing authorization code, number of packages, month of the prescription and drug cost.
For each formulation, it is possible to retrieve the information on the drug description, active ingredients, therapeutic classification (ATC) and "defined daily doses" (DDD) through record linkage between the marketing authorization code and the drugs database. In some regions, it is also

possible to link the above database (HDC, outpatient records, drug prescription) through a patient unique identification anonymous code.

Environmental data

Some of the genetic changes that are responsible for H&N cancer may be linked to the environmental exposures that damage DNA. These exposures may include substances, such as the chemicals in tobacco smoke, or radiation, such as ultraviolet rays from the sun. Asbestos exposures account for the largest percent of occupational cancer, with the greatest risks among workers who smoke [7].

Environmental contaminants may be present in the home, at work or outside. Some of these exposures, such as tobacco smoke and the sun's rays, can be avoided. Unfortunately there are also substances that are present in the air breathe, the water, the food, or the materials used during the own jobs. Differences dangerous substances were selected as the most likely carcinogens to affect human health. The 2014 edition[5] lists 56 known human carcinogens.

The identification of sources Open Data containing environmental data related with the interested regions and with the topics was conducted. All data described in Annex III are electronically available and can be extracted in any convenient format. Individual actions and public policies can help prevent or reduce the exposure to head and neck cancer causing substances in the environment: whenever possible, exposure to cancer-causing substances can be identified and stopped by using safer alternatives.

Diagnostic imaging databases/genomic databases

Diagnostic imaging databases (e.g. the forthcoming TCIA-The Cancer Imaging Archive, http://www.cancerimagingarchive.net/) or the MAASTRO images database (https://www.cancerdata.org/data?q=image_archive) are available for access and usege by the Big Data Analysis software. These datasets should be used to validate or to refine the data analysis algorithms.

The Cancer Genome Atlas (TCGA) database contains several hundreds of genome sequencing data (https://tcga-data.nci.nih.gov/tcga/tcgaCancerDetails.jsp?diseaseType=HNSC&diseaseName=Head and Neck squamous cell carcinoma) from head and neck tumor specimens. The Cancer Genomics Hub (CGHub) is a secure repository for storing, cataloging, and accessing cancer genome sequences, alignments, and mutation information from the Cancer Genome Atlas (TCGA) consortium and related projects (https://cghub.ucsc.edu/summary_stats.html).

User needs. *Access to these datasets should be required and obtained by the BD2Decide Consortium when needed, in order to achieve access to these data.*
The combination of imaging and genomic data might lead to new discoveries concerning Head and

[5] http://www.cancer.gov/about-cancer/causes-prevention/risk/substances/carcinogens

Neck Cancer phenotypes/characteristics.
The data available in these datasets might also complement retrospective data collected during the BD2Decide project, to refine the models or the Big Data analysis algorithms.

3.7 DATA EXTRACTION FOR RESEARCH AND SCIENTIFIC PUBLICATIONS

Use case 1. The clinician needs to study survival of tumors of the tongue depending on smoking habits, sex and age, and compare them with the overall OSCC tumors, to verify whether tumors of the tongue have specific prognostic characteristics. This is because the clinician has a difficult case of tumor of the tongue to be discussed in Tumor Board and he wants to produce a paper on this case.

The clinician opens BD2Decide and selects the option 'data extraction'. The system proposes a set of parameters from the data available in the system, grouped by data subset (clinical, treatment, imaging etc.). The clinician selects Tumor localization "tongue", then age range and sex and survival prediction score.

The system produces an exportable file (such as .csv, .xls, etc.) extracting the cases matching each tumor-localization sorted by age and sex and indicating the risk score and the risk predictive factors. The system produces on the screen a graph for each T-localization and sex, and a curve indicating the cumulative reoccurrence risk probability. So the clinician can see if a specific T-localization has as an average more or less survival probability as compared to the overall OSCC survival probability.

By clicking on the graph the clinician can see the data of the specific patient (i.e. age, sex, reoccurrence risk probability). The data extracted can be elaborated separately for statistical analysis and editing of scientific papers. Similar queries may be performed on different selection criteria.

Use case 2. The clinician wants to identify all patients who underwent a particular treatment (e.g. radio-chemo therapy with specific parameters) or surgical procedure (i.e. demolition and/or reconstruction) and verify the outcome of such interventions. The clinician opens BD2Decide and selects the option "data extraction" (or - as an alternative - the option "reporting"). The system proposes a set of parameters grouped by data subset (clinical, treatment, imaging etc.). The clinician selects the extraction criteria (i.e. surgical procedure) and also indicates the data to be extracted (e.g. patient sex, age, tumor staging, number of affected lymph-nodes, etc. and outcome at date from the last follow-up visit). The clinician then selects the data format for extraction. The system extracts the data in the selected format.

***User needs**. Input of data will be in logical fields, ideally through drop-down menus and not encoded. The actual data storage may be encoded. An automatic calculator should be added to the system that automatically calculates follow-up time from the entered dates. e.g. time after first histopathologic diagnosis. That would help for follow up. When publications are prepared the patient characteristics need to be grouped. Hence there should be an option to perform queries on the data combining all relevant entered parameters. The query should lead to a new database*

output file. A calculator option should be added to calculate the number of patients in specific queried groups. Local databases are usually in Access, simple calculations are made in Excel and statistical analyses are performed in SPSS. Hence there should be import and export options for Access, Excel and SPSS.

For documentation, dissemination purposes within the project and scientific needs, it should easily be possible to export all graphs, images and illustrations in an encoded form in a good quality and data format.

Use case 3. The researcher wants to identify new correlations between patient's data (multiscale) and treatment in the prognosis of Stage III and Stage IV patients. The researcher selects patients subgroups (tumor subsite, tumor stage) and eventually treatment and the system produces in output the list of relevant factors and the outcome at month 24 and 36 if available.

Use case 4. The researcher wants to cluster patients in order to understand why (in terms of contributing and additive values of the risk factors) one cluster progress better or worse than others. The researcher must be able to select groups of patients based on subjective criteria (clinical and pathology data, risk factors, genomic characteristics, HPV positive/negative). The system should score each selected factor in terms of its prognostic value for survival.

***User needs**. To create group of users corresponding to a profile that is created by the researcher and compare the results.*

Also to allow the system automatically identify the additive value of each prognostic factor for each patient selected in terms of stage, tumor subsite for the overall survival and for the final quality of life.

To estimate the value and the cost of adding certain type of information to the analysis (in this case the profile of the researcher is a PI or a Decision Maker)

The system shall allow easy and guided patients' clustering /groups selection and produce results in both pdf and anby exportable format (xls, csv,...), for further analysis or formatting for scientific publications.

4 USER NEEDS REGARDING BD2DECIDE PLATFORM OPERATIONS, PRIVACY AND SECURITY

4.1 MANAGE ACCESS AUTHORIZATIONS

Access permissions shall be guaranteed on a personal basis. UserID and Password may not be sufficient therefore more sophisticated access control must be implemented. The access authorization is also linked to data certification and signature. Only users authorized by the hospital system administrator (written authorization needed) will be granted access to BD2Decide.

No external persons - unless authorized for consultancy purposes - will have access to BD2Decide unless specifically authorized (e.g. external experts consultations) and only for Tumor Board opinions.

On a general basis, anonymized data may be viewed by authorized personnel of all BD2Decide accredited centres, for research purposes. The data shall not allow identifying the single patient. Patient's identification data will be visualized only to the hospital personnel who manage the patient.

The levels of permissions are defined above in section 2.1

4.1.1 Access

A list of BD2Decide authorized hospitals and personnel shall be maintained. During the project the Coordinator is responsible for the list of hospitals authorized for the project. After the end of the project a decision shall be taken concerning the accessibility and availability of the BD2Decide data. Each hospital PI and IT manager is responsible to authorize the clinicians for access to BD2Decide. This shall be done in line with hospital current practice and by means of an administration tool managed by the hospital department in charge.

The users should be given access credentials (no cumulative credentials allowed). Means to identify the operator who uses the system to modify (and also read) the data must be implemented, in respect of privacy regulations.

Login must be recorded and login and logout date and time must be recorded as well.

4.1.2 Automatic suspension (black screen)

It may happen that a clinician has to leave the system for urgent visits while inserting data. In this case the system must prevent un-authorized access. Therefore a clock must be set that automatically suspends the open session after 1 minute or a time to be indicated by the hospital regulations. The screen will become black and the session locked. After 30 minutes the session will be automatically closed and data not saved will be lost. As an alternative (to be verified), if possible, data shall be kept in a draft format to be clearly indicated when the system is opened again.

4.1.3 Resuming a suspended session

In case of suspension of a data entry/consultation session, within 30 minutes the clinician can resume the session without data losses. To do so the user activates the session and inserts his/her credentials again. If the credentials match the credentials inserted at login, then the session is resumed and data on the screen are presented. In case the credentials are different from previous login, a warning message must be issued:

"User: <userID> name of user already logged in." and the following options proposed:

'close current session', 'continue' or 'suspend'.

If the user decides to 'close current session', the message "data may be lost" must appear on the screen. Then a login screen is presented and the new user logs in.

If the user selects 'Continue' the login screen is presented again and the verification of credentials is done again.

If the user selects 'suspend' the session is suspended and the clock starts from the point the new user attempted the access.

4.1.4 Logout

BD2Decide will be closed manually by the user who makes logout or automatically after 30 minutes of suspension. Data not saved by the user will be lost and a message issued.

4.1.5 Automatic data saving

Automatic data saving should be prevented as a matter of data integrity and clinical value. In exceptional cases however automatic saving could be performed. One case may be when electric power stops abruptly. In this case an emergency saving should be performed in a separate area of the system. The so saved data might be used for recovery purposes.

In all cases, when emergency automatic saving is performed, a message shall be issued both before/during saving and at the time of recovery/system re-start.

4.1.6 Backup

Backups of data and logs of activities shall be kept on a daily basis to ensure that no data losses occur. Backup systems should be implemented as foreseen by each hospital policy.

4.1.7 Data preservation

Backups shall be preserved according to security procedures implemented in each hospital. For the BD2Decide system, backups shall be preserved safely and securely, in encrypted format and in securely protected intranet.

5 TECHNICAL REQUIREMENTS

The following technology requirements/ constraints and needs related to data protection, security, backup and recovery must be implemented.

5.1 DATA CERTIFICATION

Data origin must be certified (i.e. signed). In this sense the user-ID and password are not sufficient and specific signature processes must be implemented anytime data are modified. The technical team will propose solutions for this.

5.2 DATA INTEGRITY AND MAINTENANCE

Data integrity must be ensured both for BD2Decide datasets and for all the transactions of data in-out to HIS. To ensure data integrity the software must implement logging and malfunctioning detection procedures and allow for automatic data recovery. In case of recovery the system administrator must be informed and the hospital data manager involved.

Backup procedures should be put in place as required in each participating centre.

5.3 DATA PROTECTION AND SECURITY

A security layer should be implemented to avoid un-authorized access to data from third parties, according to each participating centre regulations. For the encoded DB common to all centres, adequate security protection must be ensured.

5.4 DATA COMPATIBILITY

One issue of gathering data from different hospitals consists in the establishment of a shared semantics able to uniquely identify a data independently from the name applied to it. Thus a robust semantics must be adopted for the terms and data collected in BD2Decide and used by all BD2Decide components. The technical partners need to verify the most up-to-date standards in this area at EU level and apply them to the project. ICD009/ICD10

5.5 INTEROPERABILITY WITH EXISTING LEGACY SYSTEMS

HL7 and DICOM interoperability are a standard for the developed software applications. Other interoperability requirements will be detailed by the technical analysis.

5.6 CERTIFICATIONS

The Image analysis and the radiomics software shall be certified the relevant European Authorities for usage in clinical practice. They fall under the Diagnostic Devices Class IIa category and therefore shall be approved by a Notified Body for clinical use.

For research purposes and for the execution of the project no such certification is required.

All lab equipment used for the data collection must be at least CE marked, but possibly should be certified as medical diagnostic device. This is mandatory for all exams that are part of the usual clinical practice.

The BD2Decide Big Data analysis will need further investigations regarding certification for clinical use; for the execution of the project no such certification is required.

The BD2Decide Patients Documentation System does not need certifications.

5.7 USABILITY

The BD2Decide platform components must be very user-friendly and supportive for the clinician as well as time-saving if possible. The best is, if most interactions are self-explaining and appear analogous to frequently used tools of every-day life (e.g. smartphone/tablet etc.).
Therefore the system should:

- integrate a tool for “help” to serve the software. Explanations must be understandable for “Computer dummies”
- have only few buttons that are more or less self explaining like “Apps” on an iphone
- visualize as much as possible with illustrative but clear and abstract images and use as less numerical classifications as possible
- have a simple and clearly structure surface with necessary but not irrelevant information for the first view. By continuing in different sub-datasets, it must be possible to get more detailed information in a graduated manner to understand the prognostic prediction

The BD2Decide system should not only facilitate the assessment of a tumor patient and his possible future course of illness but also facilitate the clinicians' work. It should not be realized as a time-consuming system, but as a time-saving and quality-improving software. Therefore the software should be attractive in every-day work by its prediction tools, its automated helping-tools and its easily manageable, mostly self-explaining, fast and smart appearance.

6 TECHNICAL USE CASES

In this section, a preliminary overview of the use cases that have been derived through a 1) top-down approach (i.e. derived from the user needs and scenarios defined in the previous sections of this deliverable) and a 2) bottom-up approach (i.e. derived from the building blocks of the BD2D system, as defined in *figure 5 – The BD2Decide proposed architecture* of the project DoA), is provided. The purpose of including them as part of this document is to provide the recipients with a preliminary idea of the link between the user needs and the technical and functional specifications that are need in order to design the project architecture. Therefore, the following section represent a first, not complete, but indicative idea of the main functionalities that have to be defined by the technical partners of this project. In order to have a complete list, co-design meetings with end-users and technical workshops will be carried out in the second and third quarter of project Year 1.

6.1 IMAGING FEATURES EXTRACTIONS

These use cases refer to the processing of diagnostic images (CT/MRI scans) for the automatic or semi-automatic extraction of anatomical features and for the segmentation of tumors and lymph-nodes that is required to perform the radiomics.

6.1.1 Coupled Shaped Model

Basic Flow of Use Case n°1	**Flow 1**
Title	**Coupled Shaped Model Feature extraction**
Actor	**Researchers**
Description	The coupled shape model will be used for the automatic extraction of some image features. For the BD2Decide project a new coupled shape model has to be trained and the adaptation algorithms have to be adapted. The new coupled shape model will be incorporated into the image analysis software to allow the automatic extraction of image features.
Steps	
1	The radiologist loads the medical image data into the software. This can be done from:
1a	A clinical PACS server
1b	An external storage device
2	The coupled shape model adaptation is executed.

3	The radiologist does the remaining necessary steps of the feature extraction process.
4	The software uses the coupled shape model adaptation to automatically extract certain image features.
Example	
Precondition nº	Image available
Description	The patients image data is available either online or offline
PostCondition	Data transfer
Description	Explained in 6.1.3

6.1.2 Tumor and Lymphnode analysis in MRI

Basic Flow of Use Case nº1	**Flow 1**
Title	**Tumor / Lymph node segmentation**
Actor	**Researchers**
Description	The radiologist does the segmentation of the tumor / lymph node(s)
Steps	
1	The radiologist loads the medical image data into the software. This can be done from:
1a	A clinical PACS server
1b	An external storage device
2	The radiologist uses an automatic segmentation algorithm for the segmentation of the target structure
3	If needed the radiologist manually modifies the segmentation.
4	The segmentation is saved locally and can be uploaded to the database at the end of the feature extraction process.
Example	A new patient arrives and the tumor shall be analysed.
Precondition nº	Image available
Description	The patients image data is available either online or offline
PostCondition	Data transfer
Description	Explained in 6.1.3

6.1.3 Datatransfer to database

Basic Flow of Use Case nº1	**Flow 1**
Title	**Coupled Shaped Model Feature extraction**
Actor	**Researchers**
Description	The radiologist does the feature image feature extraction process.
Steps	
1	The radiologist verifies the segmentations and extracted features.
2	The results are saved on the local storage device
3	The image features are transfer to the BD2Decide database
4	Optional: The segmentations can be send to the DICOM server
Example	
Precondition nº	Steps 6.1.1 and 6.1.2 completed and the BD2Decide database is available
Description	The medical image data is loaded and image feature extraction and tumor segmentation is already completed and a network connection to the BD2Decide database is available.
PostCondition	None
Description	

6.1.4 Radiomics features extraction from MRI and DWI

Basic Flow of Use Case nº1	**Flow 1**
Title	**MRI/DWI Feature extractor**
Actor	**Researchers**
Description	The Radiomics software allows the clinician to extract large amounts of MRI quantitative imaging features (e.g. tumor intensity, shape or texture) and to quantify tumor heterogeneity
Steps	
1	Verify existence of segmented ROI and MRI and MRI-DWI volumes and consistency

2	Researcher verifies anatomical correspondence between segmented ROI and lesion of interest
3	Features computation (e.g. intensity based, shape or texture based, wavelets) is lunched.
4	Computed features are store in the BD2Decide data repository and are made available to feed the prognostic models as part of the personalized prognostic signature.
Example	Researcher needs to characterize lesion by Radiomic features in a certain subject on a certain lesion and/or at a certain step of the treatment
Precondition nº	PR.0 All images and ROI are online and available. PR.1 ROI and lesion are aligned and checked.
Description	Preconditions for Radiomic MRI feature computation.
PostCondition	None
Description	Post conditions for Radiomic MRI feature computation.

6.1.5 Radiomics features extraction from CT and PET

Basic Flow of Use Case nº	**Flow 1**
Title	**Radiomics features extraction for a new patient**
Actor	**Clinical**
Description	The system enables the ability to extract CT and PET imaging features of a new patient
Steps	
1	Converting PET and/or CT and relative contouring structures (DICOM and DICOM-RT files) into software format
2	Visualizing graphically conversion results and allowing structure contour manipulation, if needed
3	Selecting the correct structure(GTV,CTV, etc.) where to perform the imaging feature extraction
4	Selecting different feature extraction modules for CT and/or PET imaging
5	Modifying features to calculate, imaging intensity discretization, voxel resampling size and features properties
6	Exporting imaging analysis results in excel format that are then merged with the patient's clinical and/or molecular data in the BD2Decide Clinical Decision

	Support System (CDSS)
7	Exporting screenshot of the CT and/or PET region where the analysis was performed
Example	One requires retrieving the patient's Radiomics features in order to decide together with other clinical and/or molecular data the right therapy that will be delivering. The process is involving patient imaging data and contouring structures. Patient record is updated with Radiomics features.
Precondition nº	PR.0 New patient imaging and structures data were inserted into the relevant data sources. PR.1 The relevant imaging sources are online and available. PR.2 Phrasing a query for retrieving the relevant patient's imaging data and segmentation structures.
Description	Preconditions for fetching the imaging data of the specific patient.
Post Condition	PO.0 Patient's record is updated with Radiomics features. PO.1 The relevant data sources are online and available. PO.2 The patient has a unique identifier that is available to the clinic.
Description	Post conditions for updating the patient's clinical data with Radiomics features.

Basic Flow of Use Case nº	**Flow 2**
Title	**Radiomics feature extraction for multiple patients**
Actor	**Researcher**
Description	The system enables the ability to extract imaging features from multiple patients
Steps	
1	Loading folder where patients data are stored in software format
2	Automatic structures retrieval and mapping based on structure name for multiple patients analysis
3	Selecting features extraction modules for CT and/or PET imaging
4	Modifying features to calculate, imaging intensity discretization, voxel resampling size and features properties
5	Applying the analysis on the selected structures for all the patients and exporting Radiomics analysis results in excel format for further analyses
Example	One requires retrieving Radiomics features from different patients in order to build/update/validate/test a Clinical Decision Support System. The process is

	involving patients' imaging data and contouring structures. Patients' data are updated with Radiomics features.
Precondition nº	PR.0 Patients imaging and structures data were inserted into the relevant data sources. PR.1 The relevant imaging sources are online and available. PR.2 The imaging and structures data are available in software format.
Description	Preconditions for performing feature extraction on multiple patient.
Post Condition	PO.0 Patients' data are updated with Radiomics features. PO.1 The relevant data sources are online and available.
Description	Post conditions for updating the patients' data after the analysis.

6.2 BIG DATA

Basic Flow of Use Case nº	**Flow 1**
Title	**Patient-Clinical data**
Actor	**Clinical personals**
Description	The system enables the ability to retrieve patient data from various sources.
Steps	
1	Validating the user's authentication using a unique access token (e.g. user/password login)
2	Sending and validating the client's unique identifier for accessing its files.
3	Phrasing a query for retrieving the relevant patient's data.
4	Displaying the data to the user and allowing data manipulations on it, if needed.
5	Updating the data and storing it back in the data sources for farther use.
Example	One requires retrieving the patient's basic details, medical history and imaging data in order to decide the right therapy that will be delivering. Later the medical history of the patient should be updated again. The process is involving few data sources, each containing different types of data formats and integration of the specific patient's data. This can be accomplished by inserting the data from the different sources to the system and performing one unified query using all those data sources.

Precondition nº	PR.0 Clinical data was inserted into the relevant data sources. PR.1 The relevant data sources are online and available. PR.2 The patient has a unique identifier that is available to the clinic.
Description	Preconditions for fetching the clinical data of the specific patient.
Post Condition	PO.0 Patient's clinical data will be updated for later use. PO.1 The relevant data sources are online and available. PO.2 The patient has a unique identifier that is available to the clinic.
Description	Post conditions for updating the patient's clinical data after using them.

Basic Flow of Use Case nº	**Flow 2**
Title	**Patients' data for statistic research**
Actor	**Researchers**
Description	The system enables the ability to retrieve patients' data from various sources, under set of constrains, in order to extract statistical features.
Steps	
1	Validating the user's authentication using a unique access token (e.g. user/password login)
2	Choosing a cross section of patients to be used in the statistic query.
3	Phrasing a query for retrieving the relevant data.
4	Displaying the data to the user and allowing data manipulations on it, including statistic manipulations.
5	Updating the data and storing it back in the data sources for farther use.
Example	One requires statistical data of all the patients for specific source, which have a certain behavioral aspects, smokers for instance, grouped by their ethnicity and then grouped by 5 ranges of age. The statistical data can be the percentage of symptoms relevant to the specific research, for instance.
Precondition nº	PR.0 All relevant data was inserted into the relevant data sources. PR.1 The relevant data sources are online and available. PR.2 The user is well aware of the data identifiers that are being used. PR.3 The statistic queries are valid and have been well defined.
Description	Preconditions for analysing patients' data for statistical research purposes.

Post Condition	No post conditions are required.

Basic Flow of Use Case nº	**Flow 3**
Title	**Correlation of data from different sources**
Actor	**Researchers**
Description	The system enables the ability to retrieve and correlate data from different sources.
Steps	
1	Validating the user's authentication using a unique access token (e.g. user/password login)
2	Choosing a cross section of data types to be used in the query.
3	Phrasing a query for retrieving the relevant data.
4	Displaying the data to the user and allowing data manipulations on it, including statistic manipulations and correlations parameters.
5	Updating the data and storing it back in the data sources for farther use.
Example	One requires finding the correlation between genomic sequence in different patients and the corresponding imaging features.
Precondition nº	PR.0 All relevant data was inserted into the relevant data sources. PR.1 The relevant data sources are online and available. PR.2 The user is well aware of the data identifiers that are being used. PR.3 The correlation parameters are well defined and can be translated into query. PR.4 The statistic queries are valid and have been well defined.
Description	Preconditions for analysing data correlation between data sources for research purposes.
PostCondition	No post conditions are required.

Basic Flow of Use Case nº	**Flow 4**
Title	**Patients' treatment statistics**
Actor	**Clinical personals/Researchers**
Description	The system enables the ability to retrieve patients' data and treatment for

	finding the impact of it on a set of patient.
Steps	
1	Validating the user's authentication using a unique access token (e.g. user/password login)
2	Sending and validating the client's unique identifier for accessing its files.
3	Choosing a cross section of patients to be used in the statistic query.
4	Choosing treatment types.
5	Phrasing a query for retrieving the relevant patient's data.
6	Displaying the data to the user and allowing data manipulations on it, including statistical information and correlation between treatments and patients conditions.
7	Updating the data and storing it back in the data sources for farther use.
Example	One needs the patients' clinical data along range of time, where the patients are grouped by a set of conditions and retrieve their medical state along a time range, where certain treatment was given.
Precondition nº	PR.0 All relevant data was inserted into the relevant data sources. PR.1 The relevant data sources are online and available. PR.2 The user is well aware of the data identifiers that are being used. PR.3 The patient has a unique identifier that is available to the clinic. PR.4 The statistic queries are valid and have been well defined.
Description	Preconditions for patient's treatment statistics.
PostCondition	PO.0 New treatment data should be updated in the relevant data sources. PO.1 The patient has a unique identifier that is available to the clinic.
Description	Post conditions for patient's treatment statistics.

6.3 PROGNOSTIC MODELS AND STATISTICAL METHODS

Basic Flow of Use Case nº	**Flow 1**
Title	Calculating an individual patient's prognosis under the available treatment options
Actor	**Clinicians**

Description	The system calculates a patient's prognosis under the available treatment options
Steps	
1	Validating the user's authentication using a unique access token (e.g. user/password login).
2	Sending and validating the client's unique identifier for accessing its files.
3	Retrieving the stored data for the patient (basic details, basic history, imaging and/or genomic data) and his/her treatment options.
4	Selecting from the library of prognostic models, those models that are relevant for the patient (type H&N cancer, treatment options).
5	Allowing clinician to select a subset of treatment options (optionally).
6	Allowing clinician to select from the set a suitable – as judged by patient characteristics – reference (sub)populations for which models have been recalibrated (optionally).
7	Calculating the prognosis and uncertainty for each of the selected models and treatment options.
8	Displaying the patient's prognosis – as survival probability as a function of time – and the associated uncertainty per model/treatment option to user.
9	Storing the individual patient's survival probabilities and estimates of uncertainty.
Example	A decision is needed regarding treatment of an individual patient diagnosed with oropharyngeal cancer. The data for this patient are retrieved from the system. Treatment options for this H&N cancer are radiotherapy, surgery or chemo-radiation therapy. The clinicians and patient exclude surgery as a possible treatment option. A prognosis is requested for the patient's survival after radiotherapy and chemo-radiation treatment. The models have been recalibrated using routine clinical data from patients treated in the same hospital and this recalibrated model will be used to estimate prognoses. A plot is shown on the screen with the survival probability as a function of time together with the confidence bands around the probabilities for the different treatment options (one for each prognostic model selected). The data is stored in the patient records.
Precondition nº	PR.0. Clinical data was inserted into the relevant data sources. PR.1. The relevant data sources are online and available. PR.2. The patient has a unique identifier that is available to the clinic.

	PR.3. Suitable models exist for this specific patient PR.4. Baseline risk and recalibrated baseline risks for (sub)populations have been implemented or calculated and stored
Description	Preconditions for fetching the clinical data of the specific patient.
Post Condition	PO.0. Patient's estimated prognoses are added to their records PO.1. Details of models used and reference subpopulation selected are stored. PO.2. The relevant data sources are online and available. PO.3. The patient has a unique identifier that is available to the clinic.
Description	Post conditions for being able to access the patient's estimated prognoses at a later stage for the purpose of refining, updating or consulting.

Basic Flow of Use Case nº	**Flow 2**
Title	Synthesis of an individual patient's prognoses from different models
Actor	**Clinicians**
Description	In cases in which several competing models exist for making an individual patient's prognosis, the system enables the calculation of a pooled estimate of prognosis together with its uncertainty.
Steps	
1	Validating the user's authentication using a unique access token (e.g. user/password login).
2	Sending and validating the client's unique identifier for accessing its files.
3	Requesting of prognoses of the patient by clinician.
4	Selecting of prognostic outcomes to be pooled.
5	Running the algorithm to get the pooled estimates and uncertainty.
6	Displaying the patient's pooled prognosis – as survival probability as a function of time – and the associated uncertainty per treatment option to user.
7	Storing the individual patient's pooled survival probabilities and estimates of uncertainty.
Example	Following-up on the example in Flow 1 of WP4. When estimating the prognoses of the patient under the different treatment alternatives the clinician and patient may have considered two different models, the Rietbergen model and the Ang model, each including slightly different risk-factors. The prognoses for the patient calculated with these models differ to a small extent. The clinical may

	decide to in addition calculate a pooled estimate for the prognoses of the patient synthesising the information from the two prognostic models.
Precondition nº	PR.0. Clinical data was inserted into the relevant data sources. PR.1. The relevant data sources are online and available. PR.2. The patient has a unique identifier that is available to the clinic. PR.3. Prognoses based on at least two competing prognostic models have been calculated and stored (Flow 1 of WP 4) .
Description	Preconditions for pooling prognoses from different models.
Post Condition	PO.0. Patient's estimated prognoses are added to their records. PO.1. Details of models used and reference subpopulation selected are stored. PO.2. The relevant data sources are online and available. PO.3. The patient has a unique identifier that is available to the clinic.
Description	Post conditions for being able to access the patient's pooled prognosis at a later stage for the purpose of refining, updating or consulting.

Basic Flow of Use Case nº	**Flow 3**
Title	Cost-utility analysis of decision to collect an additional source of data for improvement of prediction for an individual patient
Actor	**Clinicians**
Description	The system performs a cost-utility analysis that weighs the cost of obtaining additional data for the patient against the anticipated increase in predictive accuracy for the prognosis.
Steps	
1	Validating the user's authentication using a unique access token (e.g. user/password login).
2	Sending and validating the client's unique identifier for accessing its files.
3	Requesting of the calculated prognoses of the patient by clinician.
4	Selecting of the 'cost-utility analysis' option by the clinician.
5	Displaying the options of additional sources of data that could be obtained for this patient (e.g. pathology, genomics, imaging data).
6	Allowing the clinician to select the available modalities (pathology, genomics, imaging) and (optionally) change the costs of collecting the data.
7	Running the algorithm to calculate the expected net-gain in predictive accuracy

	for each of the different data modalities.
8	Displaying the results on the screen.
9	Storing the results of the cost-utility analysis in the system for later retrieval.
Example	A prognosis for a patient with oral cancer has been calculated on the basis of clinical an pathological data. The uncertainty in the prognosis is fairly large. The clinician wants to know whether the predictive accuracy for prognosis of the patient may be improved by collecting genomics data for this patient before making the decision to obtain this data. In other settings, the decision may be related to selection of the most informative source of additional data (e.g. imaging or genomics data).
Precondition nº	PR.1. The relevant data sources are online and available. PR.2. The patient has a unique identifier that is available to the clinic. PR.3. Prognosis has been calculated using available data and stored. PR.4. Costs of collecting each of the types of additional data (imaging, genomics) have been stored in the system. PR.5. Data has not been obtained for at least one of the modalities. PR.6. Stepwise prognostic models have been developed and implemented for this type of H&N cancer and treatment.
Description	Preconditions for performing a cost-utility analysis.
PostCondition	PO.0. Results of the cost-utility analysis are added to the patient's records PO.1. The relevant data sources are online and available. PO.2. The patient has a unique identifier that is available to the clinic.
Description	Post conditions for being able to retrieve the results of the analysis at a later stage.

Basic Flow of Use Case nº	**Flow 4**
Title	Calibration of prognostic models to (sub)population of interest
Actor	**Clinicians**
Description	To tailor predictions to the subpopulation of patients treated in a specific center or country or patients with specific characteristics
Steps	
1	Validating the user's authentication using a unique access token (e.g. user/password login)

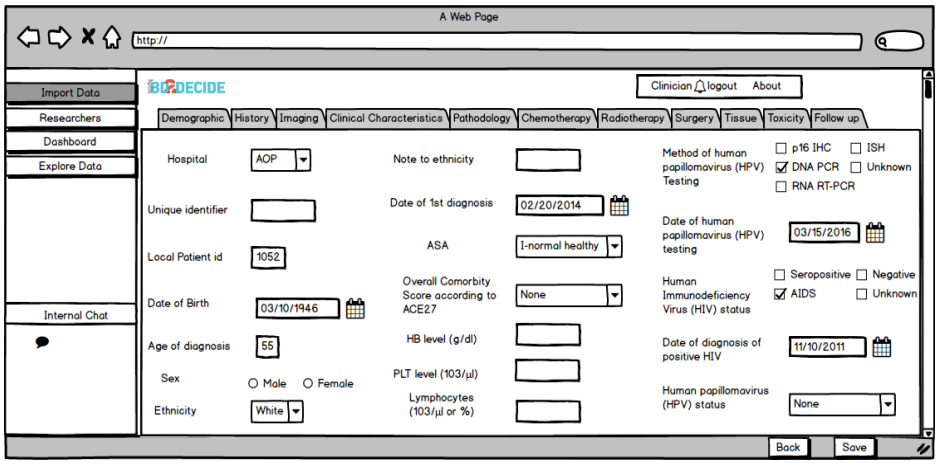
2	Choosing the models to be calibrated.
3	Choosing a subpopulation of (retrospective) patients to be used for calibration of the models.
4	Phrasing a query for retrieving the relevant data from the patients in the subpopulation.
6	Running the algorithms for updating the baseline hazards.
7	Storing the calibrated baseline hazards so that these can be applied when estimating the prognosis of future patients.
Example	A prognostic model has been derived using data from a Dutch population. In order to tailor the predictions to a different population – say an Italian population of H&N cancer patients – the baseline risk may need to be adapted. The clinician may select a representative set of Italian patients (retrospective cases) which can be used to recalibrate the models. The recalibrated baseline hazard will be stored so that it can be used for estimating prognosis of future italian H&N cancer patients.
Precondition nº	PR.0. All relevant data was inserted into the relevant data sources. PR.1. The relevant data sources are online and available. PR.2. The queries to select the relevant set of patients need to be well-defined.
Description	Preconditions for recalibration of the prognostic model to a specific (sub)population.
PostCondition	PO.0. The recalibrated baseline hazards need to be stored in order to be used in estimating prognosis of future patients. PO.1. The recalibrated baseline hazards need to be retrievable by other clinicians. PO.2. The query filed need to be stored with the baseline hazard in order to be able to verify the nature of the (sub)population selected.
Description	Post conditions for later use of the recalibrated models.

Basic Flow of Use Case nº	**Flow 5**
Title	Updating of prognostic models using routine clinical data
Actor	**Clinicians**
Description	
Steps	

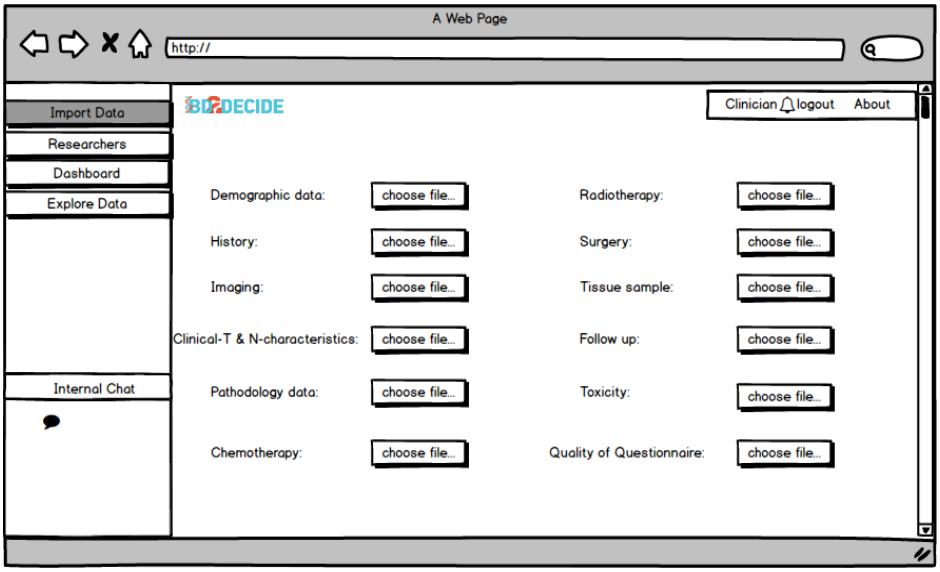
1	Validating the user's authentication using a unique access token (e.g. user/password login)
2	Choosing the models to be updated.
3	Choosing a subpopulation of (retrospective) patients to be used for updating the models.
4	Phrasing a query for retrieving the relevant data from the patients in the subpopulation.
5	Running the algorithms for updating the prognostic models (updating hazard ratio's, risk scores).
6	Storing the updated models so that these can be applied when estimating the prognosis of new patients.
Example	Routine clinical data may be used to capture geographical variations and temporal trends in characteristics of the disease and the patients treated in the hospital. A clinician may for instance choose to update the prognostic models every six months using data from the patients treated within his center.
Precondition nº	PR.0. All relevant data was inserted into the relevant data sources. PR.1. The relevant data sources are online and available. PR.2. The queries to select the relevant set of patients need to be well-defined.
Description	Preconditions for updating of the prognostic models.
PostCondition	PO.0. The updated models need to be stored in the system in order to be used in estimating prognosis of future patients. PO.1. The updated hazard ratios need to be retrievable by other clinicians. PO.2. The query filed need to be stored with the updated hazard ratios in order to be able to verify the nature of the (sub)population selected and te facilitate a similar updating procedure using a next cohort of patients.
Description	Post conditions for later use of the updated models.

6.4 PATIENT DOCUMENTATION SYSTEM.

Basic Flow of Use Case nº	**Flow 1**
Title	**Add data to PDS manually**
Actor	**Clinical personnel / clinicians**
Description	The system enables to the clinician to login and insert patient's demographic

	and clinical data into the PDS through e-CRF.
Steps	
1	Clinician logs in using authentication credentials (e.g. username/password)
2	eCRF form is presented and the clinician starts filling the patient's data (demographic and data)
3	Data from eCRF are anonymized/encrypted
4	Data are stored in the local PDS. A new record is created if patient is new, otherwise patient's record is updated.
5	Data are transferred through APIs to the big data infrastructure.
Example	
Precondition nº	PR.1: User Authentication
Description	The user must be logged in for the first time into the application
Post Condition	PO.1: Create the patient ID in PDS.
Description	The patient has a unique identifier that connects his/her to the PDS record.

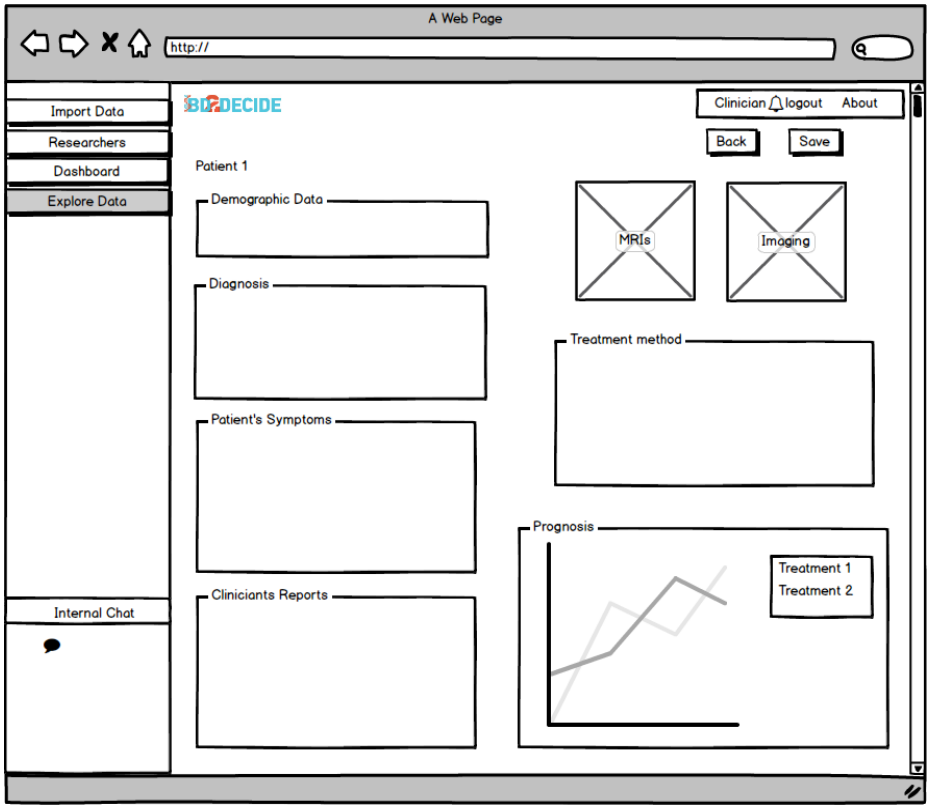
Basic Flow of Use Case nº	**Flow 2**
Title	**Add data to PDS automatically**
Actor	**Clinical IT admin (clinician personnel)**
Description	The system enables to the clinician to login and insert patient's demographic and clinical data into the PDS through a web form.
Steps	
1	Clinician logs in using authentication credentials (e.g. username/password)
2	A form is presented which indicates that the user should import files (e.g csv,jpg)
3	The demographic data from the files are anonymized/encrypted

4	Demographic and clinical data are stored in the local PDS. A new record is created if patient is new, otherwise patient's record is updated.
5	Data are transferred through APIs to the big data infrastructure.
Example	
Precondition nº	PR.1: User Authentication
Description	The user must be logged in for the first time into the application
Post Condition	PO.1: Create the patient ID in PDS.
Description	The patient has a unique identifier that connects his/her to the PDS record.

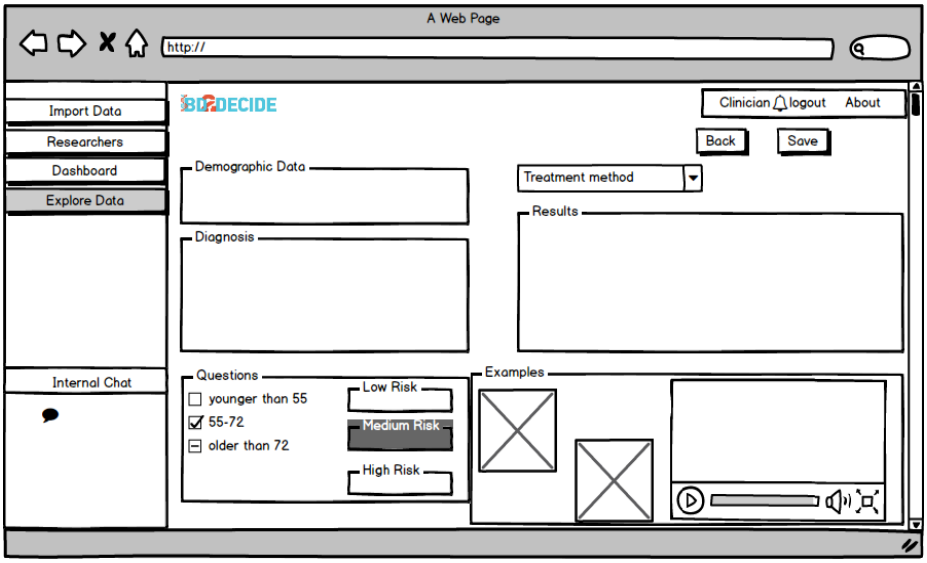
6.5 VISUAL AND PRESENTATION SUITES

6.5.1 Interactive Patient's co-Decision Aid

Basic Flow of Use Case nº	**Flow 1**
Title	**Propose treatment method**
Actor	**Clinical personnel/clinician**
Description	Clinical personnel logs in to the IPDA and selects to view patient's eCRF data and prognostic models through PDS UI.
Steps	
1	Clinician logs in using authentication credentials (e.g. username/password)
2	Select Patient

3	IPDA connects with PDS and data retrieved
4	Choose symptoms
5	Navigate to possible treatments, based on prospective analysis of patients' data.
Example	
Precondition nº	PR.1: User Authentication PR.2: Available treatment methods
Description	PR.1: The user must be logged in for the first time into the application. PR.2: The system has access to available treatment solutions.

Basic Flow of Use Case nº	**Flow 2**
Title	**Decide treatment method**
Actor	**Patient**
Description	Patient logs in to the IPDA and selects to view his profile. First he answers to some preliminary questions and after that he can decide the treatment method.
Steps	
1	Patient logs in using authentication credentials (e.g. username/password)
2	Answer to some preliminary questions

3	View possible treatment options and the proposed option of the Tumor Board
4	View the pros/cons of possible treatment options
5	Decide the treatment method according to his/her criteria
Example	
Precondition nº	PR.1: User Authentication PR.2: Available treatment methods
Description	PR.1: The user must be logged in for the first time into the application. PR.2: The system has access to available treatment solutions.
Post Condition	PO.1: Selection of treatment
Description	The patient might be able to track the proposed treatment

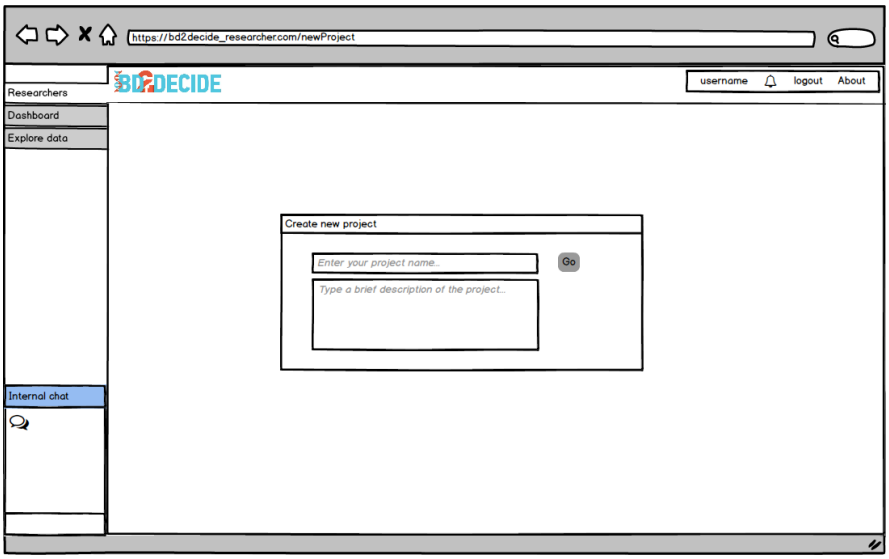
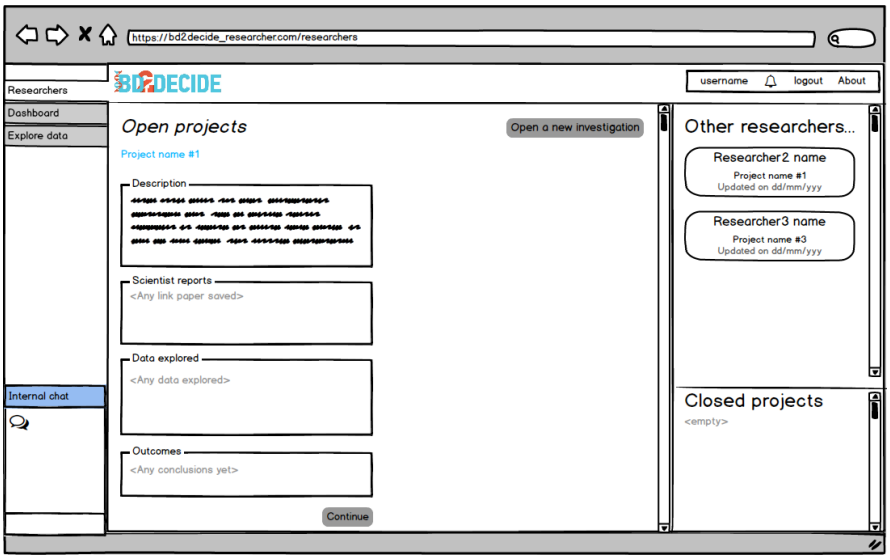
6.5.2 Clinical DSS tool suite

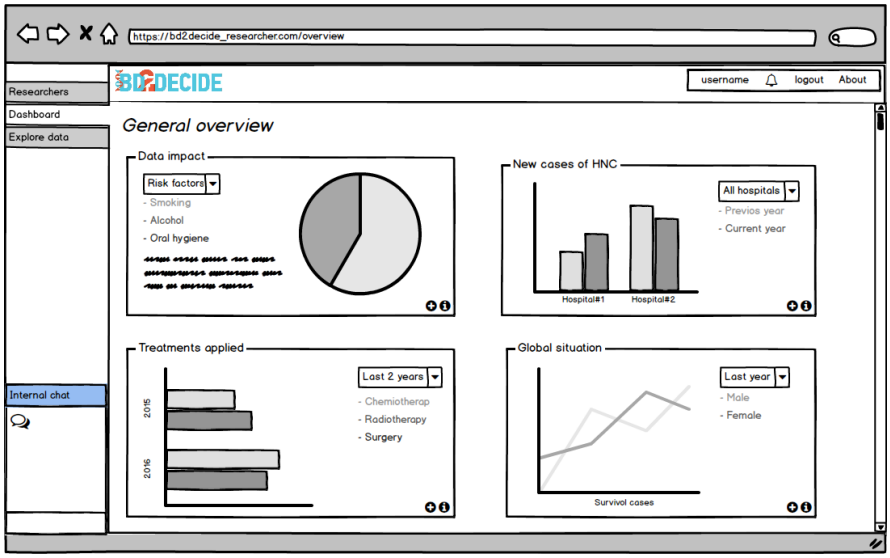
Basic Flow of Use Case nº	**Flow 1**
Title	**Clinical Decision Support System**
Actor	**Clinical personnel / clinician**
Description	Clinical personnel log into to the clinical DSS and select to view patient's info and prognostic models through CDSS UI. According to this they can make changes to the 3D model or data and see how the model changes, the therapy or other important information to personalized treatment.
Steps	
1	Clinician A and Clinician B of different specialties log in using authentication

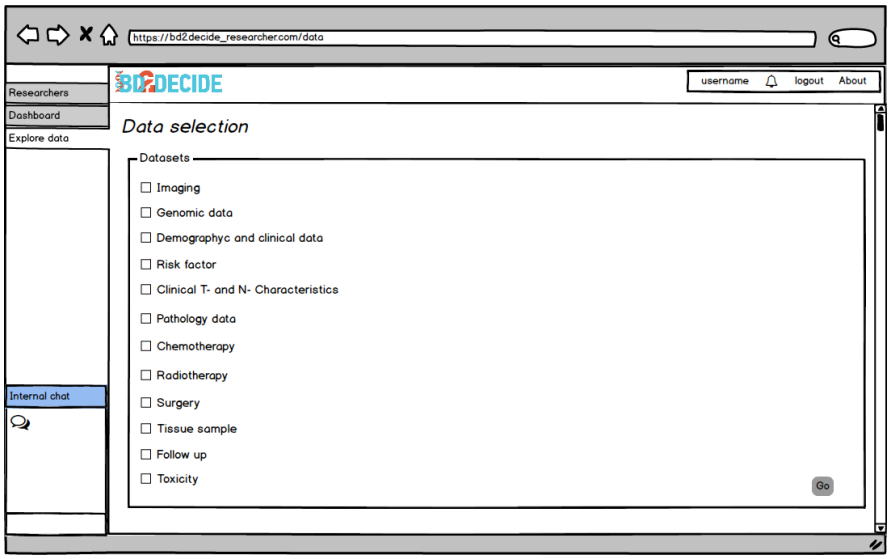
	credentials (e.g. username/password).
2	Select Patient.
3	Clinical DSS connects to the big data infrastructure and retrieves the personalized data.
4	View visualized model and the report with prognostic and/or predictive information regarding the patient and treatment options.
5	Present the information or follow-up the results of the patient's therapy.
6	The clinicians may use messaging functions to discuss on potential therapies and how they affect the clinical status of the patient from their specialty perspective.
7	The clinicians may select between different treatments and/or clinical data and assess the evolution of the patients' prognostic model in time.
8	Update data into the big data infrastructure.
Example	
Precondition nº	PR.1: Clinicians' authentication
Description	The user must be logged in for the first time into the application

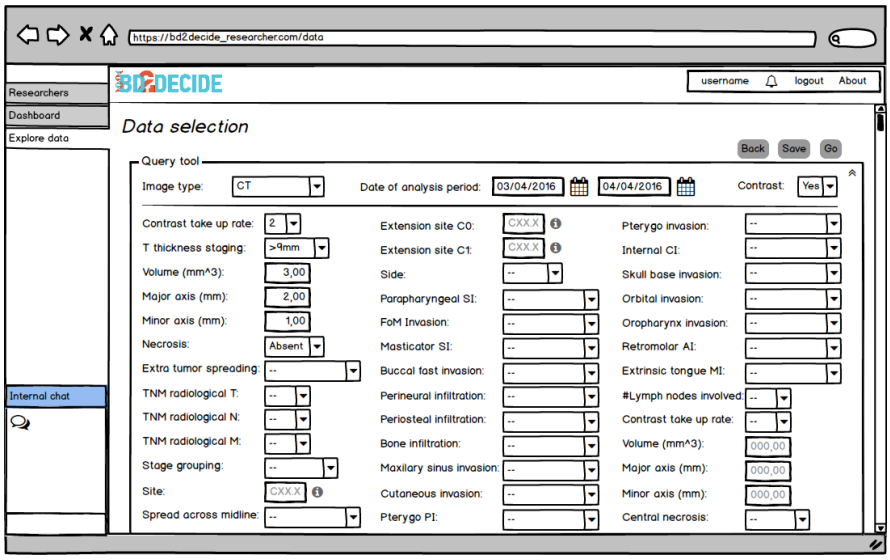
6.5.3 Researcher Scenario

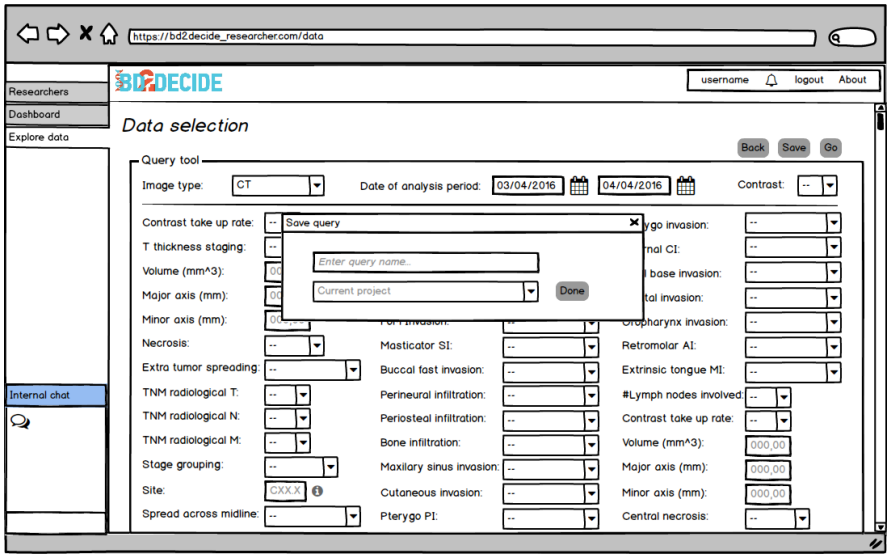
Basic Flow of Use Case nº	**Flow 1**
Title	**Create researcher profile**
Actor	**Researcher and administrator**
Description	The first time a user uses the Research tool, a new project session must be

	created. That allows the storage of all researches and queries requests
Steps	
1	Enter the name and a brief description of the project and start the investigation
Example	
Precondition nº	PR.0
Description	The user must be logged in for the first time into the application
PostCondition	PO.0
Description	The investigation starts using the tool:
	

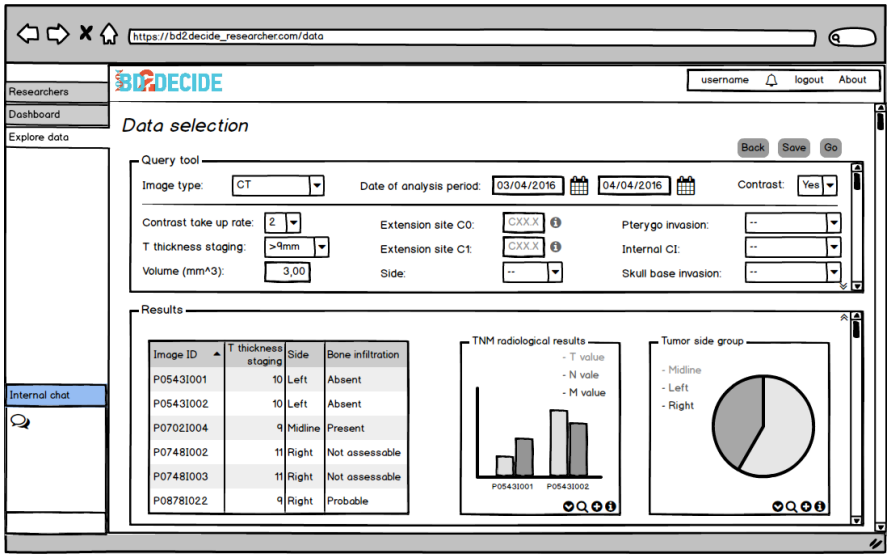
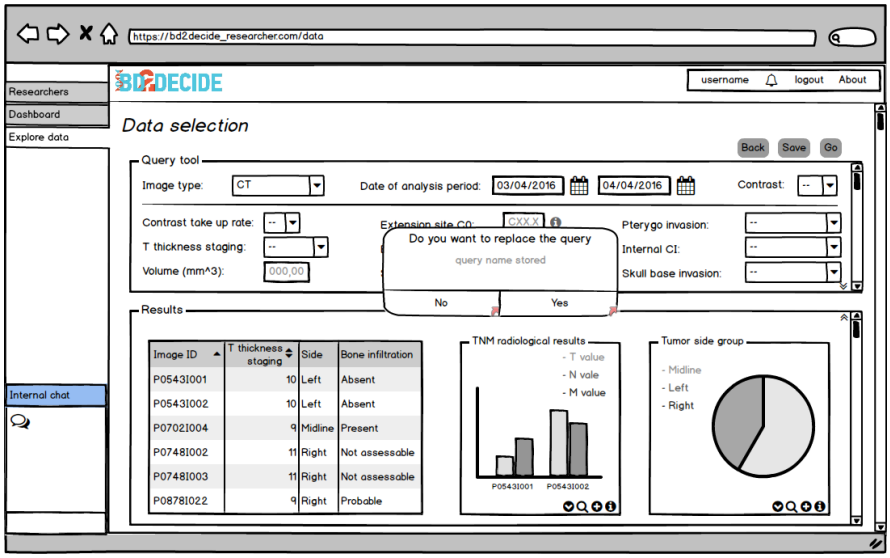
Basic Flow of Use Case nº	**2**
Title	**General overview**
Actor	**Researcher and administrator**
Description	Dynamic general overview of HNC worldwide data
Steps	
1	By selecting the *Dashboard tab*, a set of graphics trends display the general information about HNC status, based on the datasets available
2	By changing the common box, it is possible to change the data represented in each section
3	Clicking the icon, a brief description of the represented data appears
4	Clicking the icon, the represented data appears in the *Explore data tab* in order to allow the research interacting with the tool
Example	
Precondition nº	PR.0
Description	The user is logged in the application
PostCondition	PO.0
Description	The user is able to work with the selected data or carry out further investigation by clicking the *Explore data tab* (go to UC#4)

Basic Flow of Use Case nº	**3**
Title	**Explore data – Data selection (Datasets)**
Actor	**Researcher and administrator**
Description	Dataset options visualization (Based on CRF inputs)
Steps	
1	In the *Explore data tab* it is possible to select those datasets the user choose to work with
2	By selecting Imaging dataset, a filter field for imaging data appears
Example	
Precondition nº	PR.0
Description	a) Click the *Explore data tab* b) Click the ⊕ icon from *Dashboard tab* c) Click the *Continue button* from *Researchers tab*
PostCondition	PO.0
Description	A filter field option appears with the imaging dataset selected (UC#5)

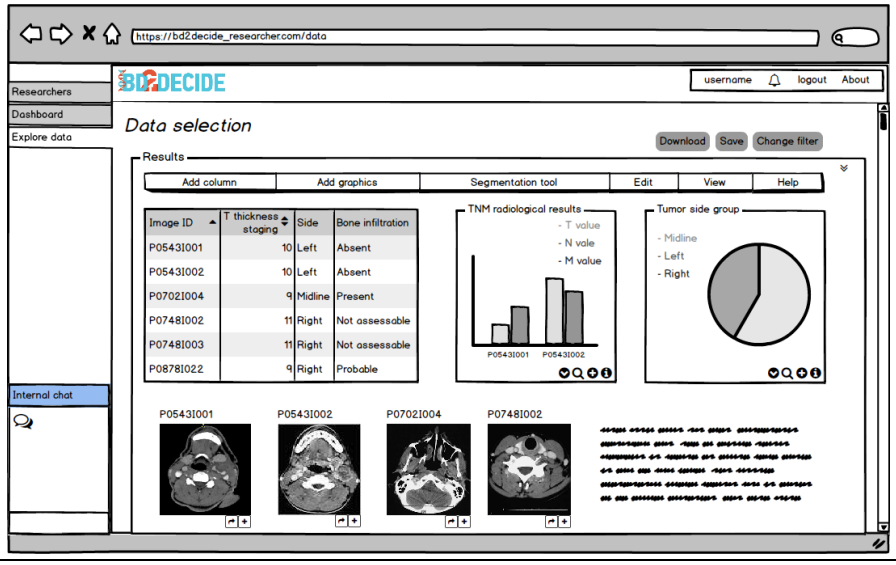
Basic Flow of Use Case nº	**4**
Title	**Explore data – Data selection (Query options)**
Actor	**Researcher and administrator**
Description	Query tool options (Based on CRF inputs)
Steps	
1	Select image type (CT, for instance)
2	Select additional conditions (e.g.: whether the image should include contrast or not)
3	Click the *Go button* to visualize the results
Example	
Precondition nº	PR.0
Description	Dataset should be selected (UC#4)
PostCondition	PO.0
Description	Data from the query created is shown below the *Query tool area* (UC#6)
PostCondition	PO.1
Description	The query can be saved, in order to be reused at a later stage, by clicking the *Save button*



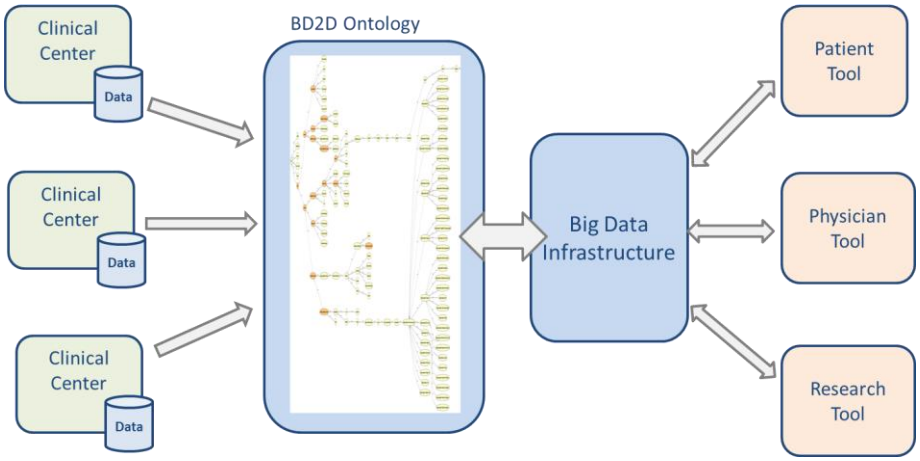
Basic Flow of Use Case nº	**5**
Title	**Explore data – Data selection (Visualization results)**
Actor	**Researcher and administrator**
Description	Visual analytics tool after selecting the desired data
Steps	
1	Visual analytics tool shows some tables and graphics according to the filtered fields selected
2	Clicking the ℹ icon, a brief description of the data appears
3	Clicking the ⊕ icon, the data appears in a new window table
4	Clicking the 🔍 icon, the papers related to the data are searched
5	Clicking the ⬇ icon, the data tables and the graphics are downloaded in different formats (e.g. pdf, xsl…)

Example	
Precondition nº	PR.0
Description	The query tool form should be filled
PostCondition	PO.0
Description	If the *Query tool area* is modified, the new query could be saved in another query link or overwrite the existing one by clicking the *Save button* 
PostCondition	PO.1
Description	With the purpose of working with the data resulting from the current query, the *Results area* will be expanded by clicking the ⌃⌃ icon (UC#7)

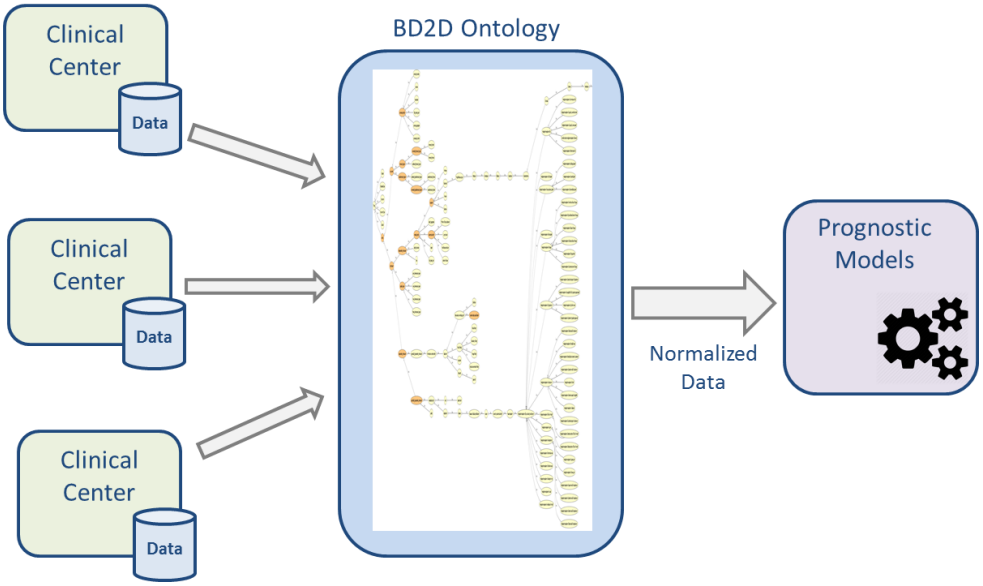
Basic Flow of Use Case nº	**6**
Title	**Explore data – Data selection (Visual analytics section)**

Actor	**Researcher and administrator**
Description	Visual analytics tool
Steps	
1	Visual analytics tool shows a set of tables and graphics, according to the filtered fields selected, including images and clinical metadata visualization
2	It is possible to add additional information by clicking the *Add column section* in the *toolbar*
3	To add more graphics, click *Add graphics* in the *toolbar*
4	It is possible to process one or more images, selecting the desired images and then, the *Segmentation tool section* in the *toolbar*
5	Clicking the icon, zoom is applied to the image
6	Clicking the icon, the image is opened in the *Segmentation tool*
7	By selecting the *Edit section* in the *toolbar* it is possible to make changes in the *Result area* By selecting the *View section* in the *toolbar* it is possible to move the object in the desired visualization
8	The *Help section* gives information about the functionality of the *Results area*
9	*Download button* allows to export the *Results area* in multiple formats (e.g.: pdf and xsl)
10	
Example	
Precondition nº	PR.0
Description	Query form should be filled

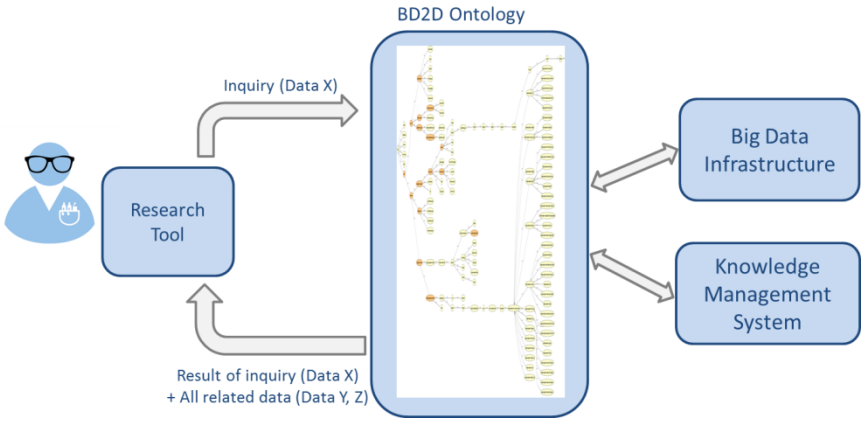
6.6 ONTOLOGIES AND KNOWLEDGE MANAGEMENT SYSTEM

Basic Flow of Use Case nº	**Flow 1**
Title	**Ontology – Interoperability**
Actor	**Clinicians Patients Researchers**
Description	Usage of the ontology for a common and data structure definition. This allows the interoperability between different systems.
Steps	
1	The user researcher opens the analytics research tool.
2	The researcher work with data related with patient clinical data.
3	The data warehouse provides the research with data from the Patient class of the ontology (Virtual Patient).
Example	

Basic Flow of Use Case nº	**Flow 2**
Title	**Ontology – Prognostic Models**
Actor	
Description	The ontology provides the infrastructure to normalize the information gathered from different clinical centres. The data defined and structured in the ontology will be used to develop, to train and validation of the prognostic models.
Steps	
1	The data from different clinical centers is structured according to the rules of the

	ontology.
2	The data are normalized.
3	The data are used by the prognostic models.
Example	

Basic Flow of Use Case nº	**Flow 3**
Title	**Ontology – Semantic**
Actor	**Researcher using the Research tool.**
Description	The ontology provides support to research through search assistance. That means the ontology allows a researcher make an inquiry related with scientific specific data, and receive structured information about all data related with this concept. This improves the research quality of the user.
Steps	
1	The user researcher opens the analytics research tool.
2	The researcher makes an inquiry about specific data (data from patients, data from research studies).
3	Using the structure and definition of the ontology, the tool improves the researcher's search through a better result. The analytics research tool provides the researcher with all data related with the specific data of his inquiry.

Example	 BD2D Ontology Inquiry (Data X) Research Tool Big Data Infrastructure Knowledge Management System Result of inquiry (Data X) + All related data (Data Y, Z)

ANNEX I - IDENTIFICATION OF H&N CANCER LOCI IN ANATOMIC IMAGES

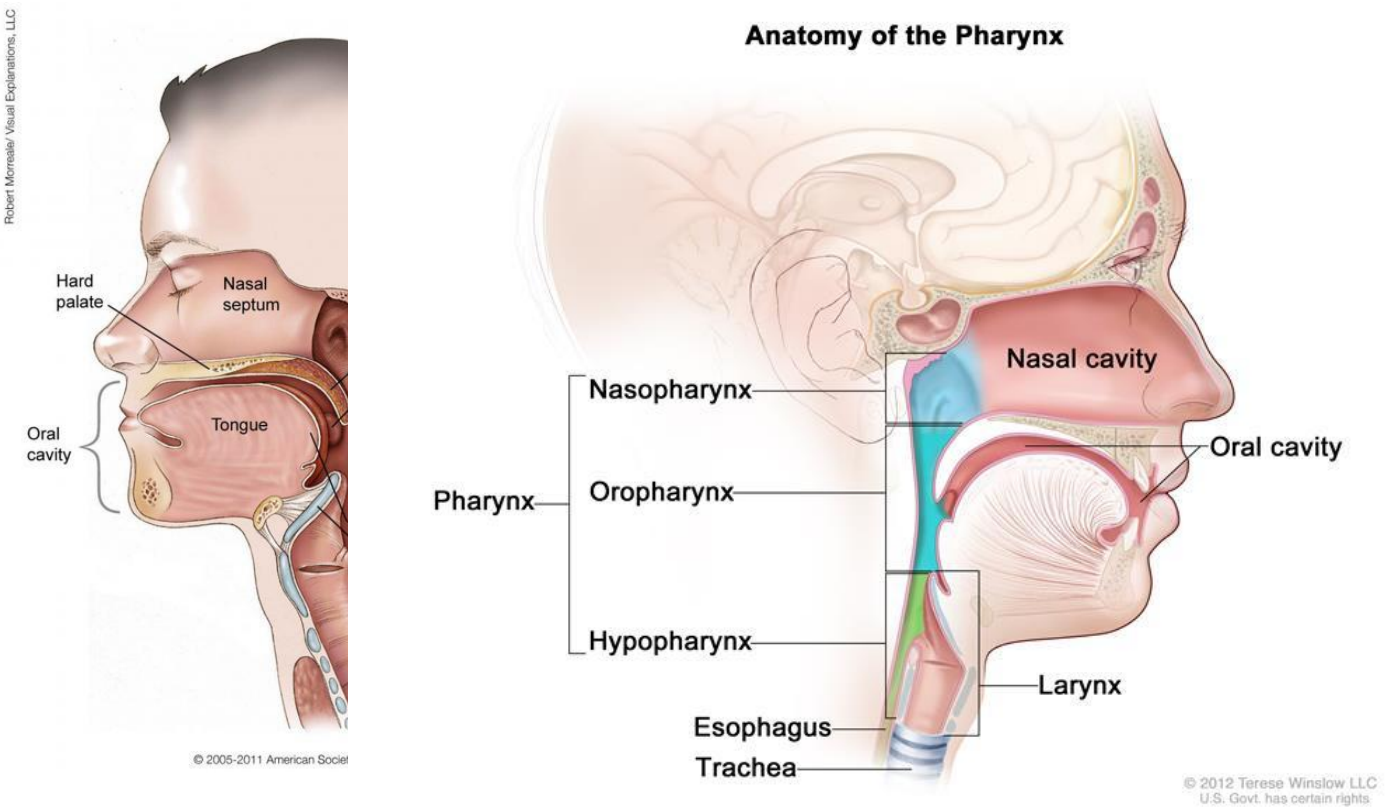


Figure 7. Anatomy of H&N Cancers

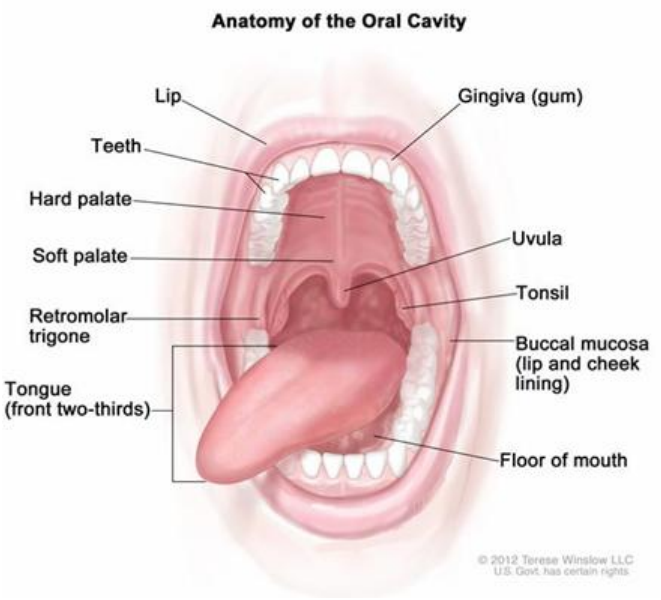
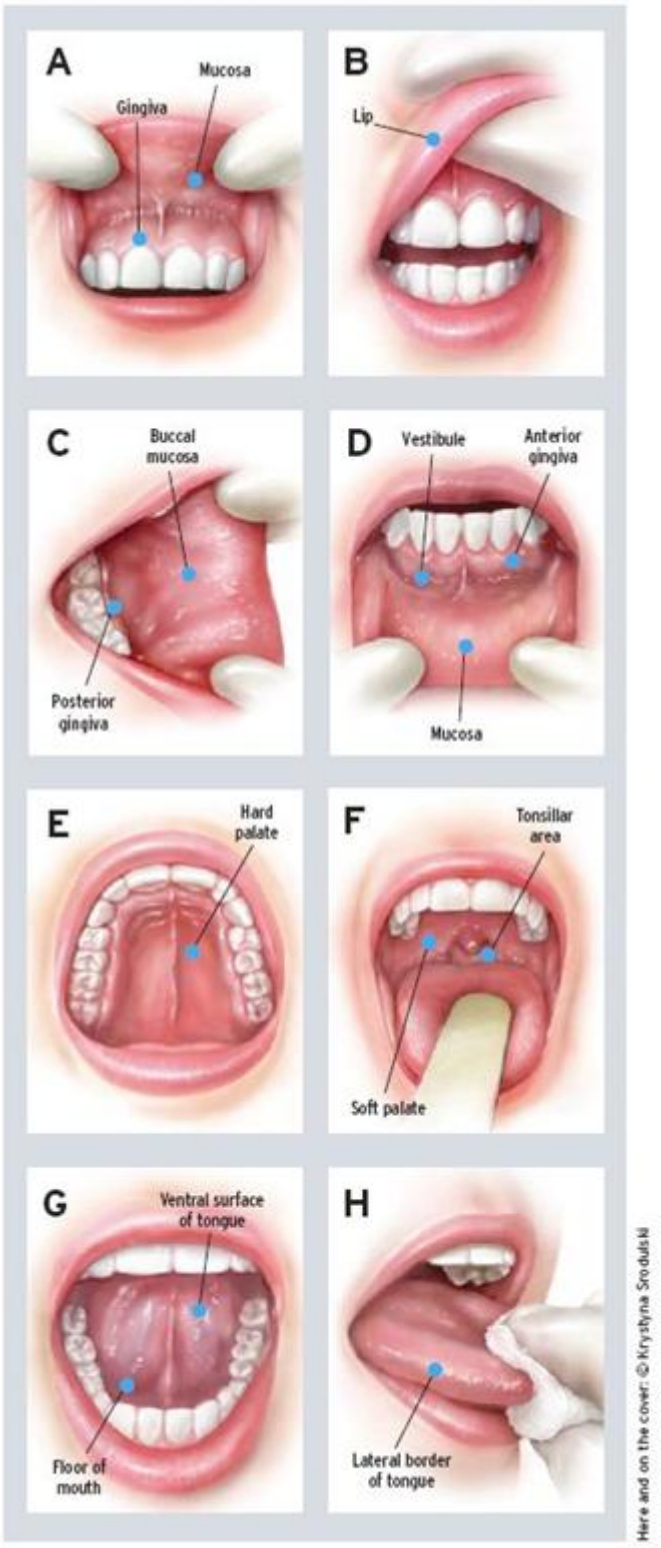


Figure 8. Anatomy of the oral cavity



- Oral Tongue
- Tongue Base
- Buccal Mucosa
- Floor of the Mouth
- Upper Gingiva
- Lower Gingiva
- Retromolar Trigone
- Hard Palate
- Soft Palate
- Tonsillar Fossa
- Superior Alveolar Ridge
- Inferior Alveolar Ridge

Figure 9. Oral cavity cancer localization

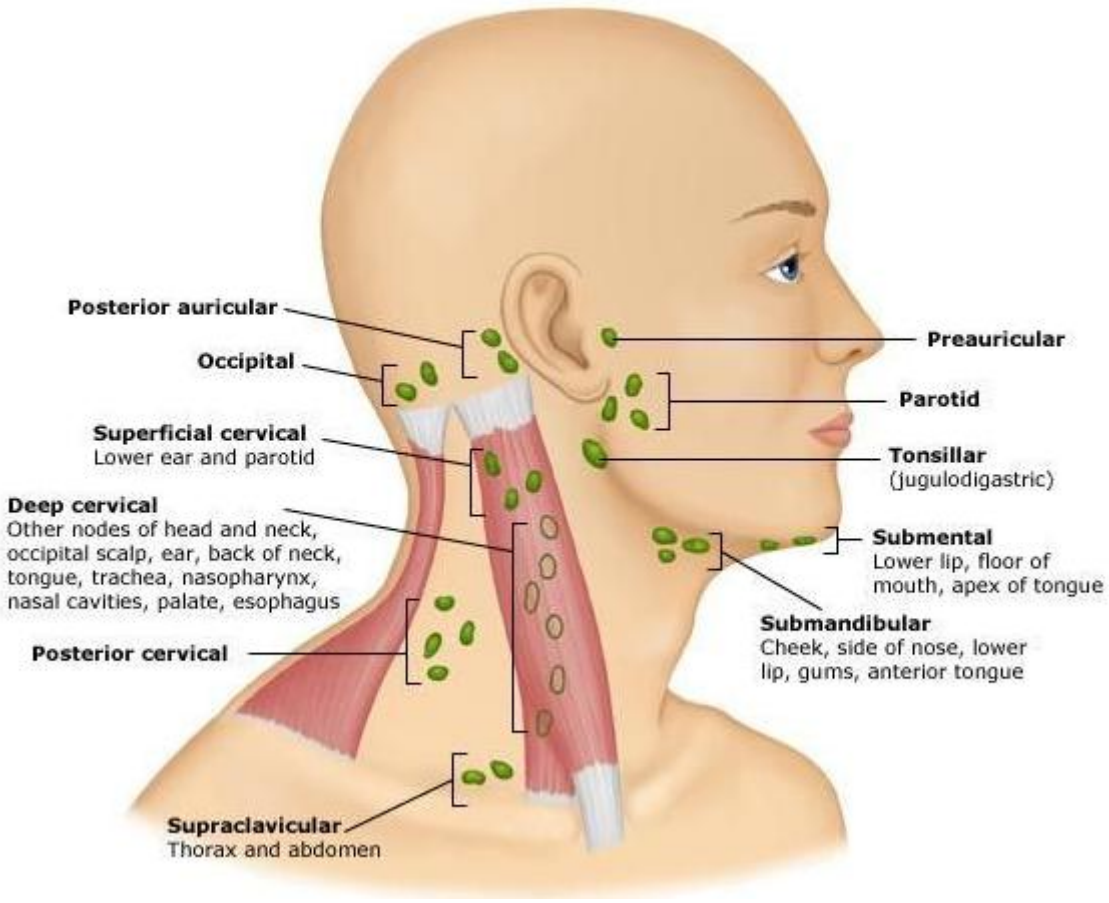


Figure 10. Anatomy of H&N Lymphnodes

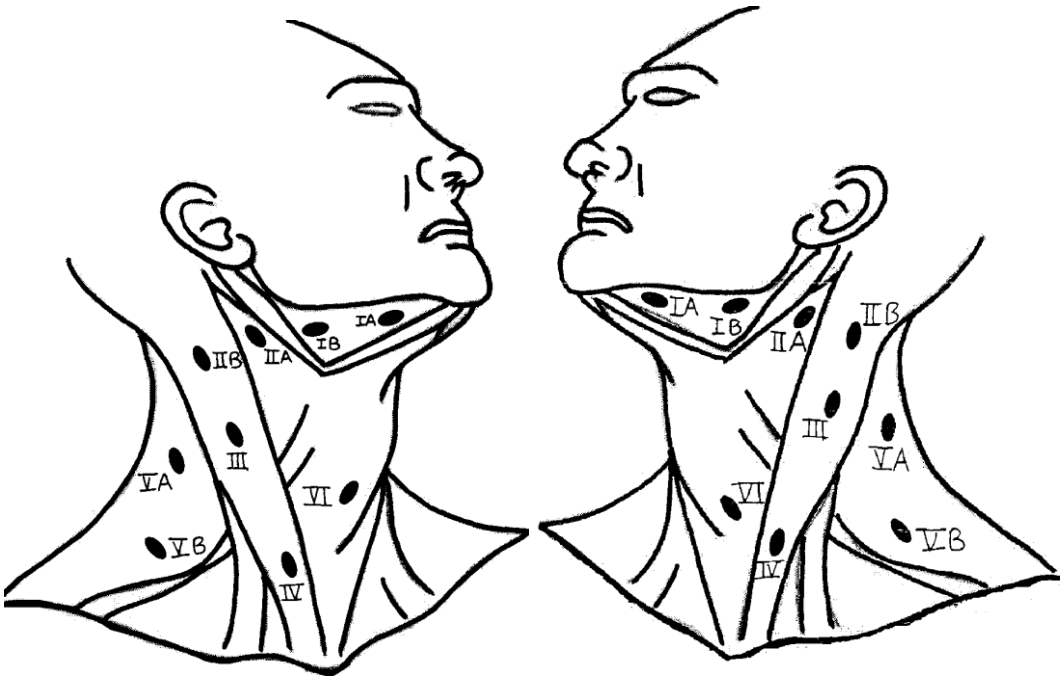


Figure 11. Lymph-nodes in neck - different levels

ANNEX II - PATIENT'S DATA (BD2DECIDE PATIENTS ELECTRONIC HEALTH RECORD)

Summary

Item 1 - Patient selection

Item 2 - Demographic & Clinical data

Item 3- Risk factor

Item 4 - Clinical T- and N-Characteristics

Item 5 – Imaging

Item 6 - Pathology data

Item 7 - Chemotherapy

Item 8 – Radiotherapy

Item 9 - Surgery

Item 10 - Tissue Sample

Item 11 - Follow up

Item 12 – Toxicity

Item 13 – Quality of Life questionnaires

NOTES:

1) Types of format:

- Radio-Buttons: allows the user to choose only one of a predefined set of options.
- Check boxes: A check box permits the user to make multiple selections from a number of options.
- Drop-Down List: A drop-down list allows the user to choose one value from a list.
- Select Lists: A list box allows the user to select one or more items from a list contained within a static, multiple line text box.

2)) data selected from drop down lists marked with an *:

The items of the drop down list should be managed in separate tables for which a data entry and data management should be provided to the system administrator, in order to allow management of the list items (e.g. add/remove items).

Item 1 - Patient selection checklist

Data	Value	Format	Notes	M
Does the patient or next-of-kin (for deceased patient) give consent to participate in the study?	1. Yes - Patient 2. Yes - Next of Kin/Legal Representative 3. Not Applicable	Drop-Down list	• for patients who are deceased at study entry • No Informed Consent is required per local/national regulations	M
If applicable please provide date of ICF for data collection	Date	DD/MM/YYYY		M
Did the patient meet all eligibility criteria? • Stage III or IVa or IVb SCCHN? • Patient treated with curative intent? • Availability 2. of CT scan or MRI images? 3. Macrodissected slides? 4. 3D or IMRT if applicable?	1. No 2. Yes	Radio-Button		M
If No please report the reason of screen failure	text	Text-area	Dependent from choice 1. in " Did the patient meet all eligibility criteria "	

Item 2 - Demographic & Clinical data

Data	Value	Format	Notes	M[6]
Patient ID	text	Text-box		M
Hospital	1. AOP 2. VUMC 3. INT 4. UDUS 5. MAASTRO 6. ULM 7. GECCP 8. MUV 9. OTHER	Drop down list *	Automatic from the login	M
Unique identifier	text	Text-box		
Local Patient ID	text	Text-box	From Hospital records	M
Date of Birth (DoB)	date	DD/MM/YYYY		M
Age at diagnosis	quantitative	000	Automatic calculation: (DoD) - (DoB)	M
Sex	1. Male 2. Female	Radio-Button		M
Ethnicity	White Black or African Asian Other, please specify	Drop-Down list *	If the selection is made by a drop down list, an extra box to specify "Ethnicity" not contemplated in the choice above is needed	M
Note to ethnicity	text	Text-box	Dependent from choice 4. in "Ethnicity"	
Date of first diagnosis (histology) (DoD)	date	DD/MM/YYYY		M
ASA	I - normal healthy patient II - patient with mild systemic disease III - patient with severe systemic disease IV - patient with severe systemic disease that is a constant threat to life V - moribund patient who is not expected to survive without the operation	Drop-Down list	https://www.asahq.org/resources/clinical-information/asa-physical-status-classification-system	M

[6] M= Mandatory, A=Automatically calculated

Data	Value	Format	Notes	M[6]
	Not Available			
Overall Comorbity Score according to ACE27	0 - None 1 - Mild 2 - Moderate 3 - Severe 9 - Unknown	Drop-Down list *	http://siog.org/files/public/ace27.pdf The ACE27 table should be shown on demand (e.g. put a clickable)	M
HB level (g/dl)	quantitative	00,0	At time of diagnosis; Normal ranges: male 13-18; female 12-16	M
PLT level (10^3/µl)	quantitative	0000,0	At time of diagnosis; Normal ranges: 150-450	M
Lymphocytes (10^3/µl or %)	quantitative	0000,0	At time of diagnosis; Normal ranges: 0,9-5,2 or 20-45%	M
Human papillomavirus (HPV) status	1. Positive 2. Negative 3. Unknown	Drop-Down list		
Method of human papillomavirus (HPV) Testing	1. p16 IHC 2. DNA PCR 3. RNA RT-PCR 4. ISH 5. Unknown	Check-box		
Date of human papillomavirus (HPV) testing	date	DD/MM/YYYY		
Human Immunodeficiency Virus (HIV) status	1. Seropositive 2. AIDS 3. Negative 4. Unknown			
Date of diagnosis of positive HIV	date	DD/MM/YYYY	Dependent from choice 1. and 2. in "Human Immunodeficiency Virus (HIV) status"	
Notes	text	Text-box		

Item 3- Risk factors

Data	Value	Format	Notes	M
Familial history of malignancies	1. No 2. Yes	Radio-Button		M
Note to Familial history of malignancies	text	Text-box	Dependent from choice 2. in "Familial history of malignancies"	
Smoker	1. Current 2. Former 3. Never 4. Unknown	Drop-Down list		M
Smoking habits (mainly)	1. Cigarettes 2. Cigar/pipes 3. Betel quid 4. Smokeless (spit) Tobacco	Drop-Down list	Dependent, if Smoker= 1 or =2	M
Packs smoked per Day	quantitative	00,0	Dependent if Smoker= 1 or =2 1pack = numbers per Day/20 per cigarettes or /4 cigars	M
Years as a smoker	quantitative	00,0	Dependent if Smoker= 1 or =2	M
Pack years	quantitative	00,0	Dependent if Smoker= 1 or =2 Calculation = (packs smoked per day) × (years as a smoker) packs smoked per day: 1pack = numbers per Day/20 for cigarettes or /4 cigars	M
Alcohol	1. Current 2. Former 3. Never 4. Unknown	Drop-Down list		M
Number of alcohol units per Day (liters)	quantitative	00,0	Dependent if alcohol = 1, 2	M
History of alcohol dependence	1. No 2. Yes 3. Unknown	Drop-Down list	Dependent if alcohol = 1, 2	M
Oral hygiene	1. Good	Drop-Down list		M

Data	Value	Format	Notes	M
	2. Poor 3. Bad			
Additional precancerous lesion	1. No 2. Leukoplakia 3. Lichen ruben planus 4. Erythroplakia 5. Oral submucous fibrosis	Drop-Down list *		M
Location of additional precancerous lesion	ICD10 definition	Drop-Down list * the codes should be shown for selection (ICD 10 TABLE)	Dependent if ADDITIONAL PRECANCEROUS LESIONS ≠ 1 Refers Addendum A A conversion table for ICD-O3 coding is needed to match tumor regisstry data	
Notes	text	Text-box		

Item 4 - Clinical T- and N-Characteristics

Data	Value	Format	Notes	M
TNM cT	1. T1 2. T2 3. T3 4. T4a 5. T4b	Drop-Down list *		M
TNM cN	1. Nx 2. N0 3. N1 4. N2a 5. N2b 6. N2c 7. N3	Drop-Down list *	If N3 provide data on tissue invasion (see "Tissue invasion for cN3")	M
Tissue invasion for cN3	1. Skin 2. Muscle 3. Cranial nerve XI 4. Brachial plexus 5. Parotid gland	Select list *	Dependent from choice 7. in "TNM cN"	M

Data	Value	Format	Notes	M
Tumor Region	1. Oral Cavity 2. Oropharynx 3. Hypopharynx 4. Larynx	Select list *		M
Anatomical Tumor Location	ICD10 definition	Select list *	Ref. Addendum A	M
Laterality	1. Left 2. Right 3. Midline	Drop-Down list		M
N Side	1. Ipsilateral 2. Contralateral 3. Bilateral	Drop-Down list	Dependent if cN ≠ 1 and cN≠2	
Node levels right	1. IA 2. IB 3. IIA 4. IIB 5. III 6. IV 7. VA 8. VB 9. VI	Select List *	Dependent if cN ≠ 1 and cN≠2	
Node levels left	1. IA 2. IB 3. IIA 4. IIB 5. III 6. IV 7. VA 8. VB 9. VI	Select List *	More than 1 choice Dependent if cN ≠ 1 and cN≠2	
Stage at diagnosis	1. Stage 0 2. Stage I 3. Stage II 4. Stage III 5. Stage IVA 6. Stage IVB 7. Stage IVC	Drop-Down list *	Show the whole dataset for completeness	M
Notes				

Item 5 – Imaging

(automatically extracted from Fraunhofer's Image analysis software and Radiomics software)

Data	Value	Format	Notes	M
Case number (Image ID)	text	Text-box		M
Date of analysis	date	DD/MM/YYYY		M
Image type	1. CT 2. MRI 3. PET/PET-CT 4. Ultrasound	Drop-Down List	Each Image analysis (CT or MRI or PET/PET-CT or ultrasound) should have a dedicated dataset section starting from "Tumor" because is a single analysis of segmentation	M
Contrast	1. No 2. Yes	Radio-Button		M
Tumor Imaging data				
Contrast take up rate	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	Drop-Down List	Edge quality: 0 = irregular 10 = sharp borders	
T Thickness Staging	1. 3mm 2. 3mm ≤ 9mm 3. > 9mm	Drop-Down List		M
Volume (mm^3)	quantitative	000,00		M
Major axis (mm)	quantitative	000,00		M
Minor axis (mm)	quantitative	000,00		M
Necrosis	1. Absent 2. Present	Drop-Down List		M
Extra tumor spreading	1. Absent 2. Probable 3. Present 4. Not assessable	Drop-Down List		M
TNM radiological T	1. T1 2. T2 3. T3 4. T4a 5. T4b	Drop-Down List *		M
TNM radiological N	1. Nx 2. N0 3. N1 4. N2a 5. N2b 6. N2c 7. N3	Drop-Down List *		M

Data	Value	Format	Notes	M
TNM radiological M	1. Mx 2. M0 3. M1	Drop-Down List *		M
Stage grouping	1. Stage 0 2. Stage I 3. Stage II 4. Stage III 5. Stage IVA 6. Stage IVB 7. Stage IVC	Drop-Down List *		A
Site	ICD10 definition	Drop-Down list *	Refers Addendum A	M
Spread across midline	1. Absent 2. Probable 3. Present 4. Not assessable	Drop-Down List *		
Extension Site C0	ICD10 definition	Drop-Down list *	Refers Addendum A	M
Extension site C1	ICD10 definition	Drop-Down list *	Refers Addendum A	M
Side	1. Left 2. Right 3. Midline	Select-List		M
Parapharyngeal space invasion	1. Absent 2. Probable 3. Present 4. Not assessable	Drop-Down List		M
Floor of the mouth invasion	1. Absent 2. Probable 3. Present 4. Not assessable	Drop-Down List		M
Masticator space invasion	1. Absent 2. Probable 3. Present 4. Not assessable	Drop-Down List		M
Buccal fat invasion	1. Absent 2. Probable 3. Present 4. Not assessable	Drop-Down List		M
Perineural infiltration	1. Absent 2. Probable 3. Present 4. Not assessable	Drop-Down List	ONLY FOR MRI	M
Periosteal infiltration	1. Absent 2. Probable 3. Present 4. Not assessable	Drop-Down List		M
Bone infiltration	1. Absent	Drop-Down List		M

Data	Value	Format	Notes	M
	2. Probable 3. Present 4. Not assessable			
Maxillary sinus invasion	1. Absent 2. Probable 3. Present 4. Not assessable	Drop-Down List		M
Cutaneous invasion	1. Absent 2. Probable 3. Present 4. Not assessable	Drop-Down List		M
Pterygo palatine invasion	1. Absent 2. Probable 3. Present 4. Not assessable	Drop-Down List		M
Pterygo invasion	1. Absent 2. Probable 3. Present 4. Not assessable	Drop-Down List		M
Internal carotid invasion	1. Absent 2. Probable 3. Present 4. Not assessable	Drop-Down List		M
Skull base invasion	1. Absent 2. Probable 3. Present 4. Not assessable	Drop-Down List		M
Orbital invasion	1. Absent 2. Probable 3. Present 4. Not assessable	Drop-Down List		M
Oropharynx invasion	1. Absent 2. Probable 3. Present 4. Not assessable	Drop-Down List		M
Retromolar area invasion	1. Absent 2. Probable 3. Present 4. Not assessable	Drop-Down List		M
Extrinsic tongue musculature invasion	1. Absent 2. Probable 3. Present 4. Not assessable	Drop-Down List		M
Lymph nodes (Imaging data				
Number of Lymph nodes	1, 2, 3, 4, 5, 6, 7,	Drop-Down List	For the 3 most	M

Data	Value	Format	Notes	M
involved	8, 9		important lymph nodes involved indicated with value from 1 to 9 the "Lymph nodes" section shall be repeated to collect the correspondent dataset	
Contrast take up rate	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	Drop-Down List	Edge quality: 0 = irregular 10 = sharp borders	M
Volume (mm^3)	quantitative	000,00		M
Major axis (mm)	quantitative	000,00		M
Minor axis (mm)	quantitative	000,00		M
Central necrosis	1. Absent 2. Present	Drop-Down List		M
Necrosis	1. Absent 2. Present	Drop-Down List		M
Extra nodal spreading	1. Absent 2. Probable 3. Present 4. Not assessable	Drop-Down List		M
Extra nodal spreading: irregular margin edges	1. Absent 2. Present	Drop-Down List	Alteration of the surrounding fat / irregular margins edges	M
Shape deviation	quantitative	000,00		
Bone infiltration	1. Absent 2. Probable 3. Present 4. Not assessable	Drop-Down List		M
Carotid infiltration	1. Absent 2. Probable 3. Present 4. Not assessable	Drop-Down List		M
Cutaneous invasion (skin)	1. Absent 2. Probable 3. Present 4. Not assessable	Drop-Down List		M
Side	1. Left 2. Right 3. Midline	Drop-Down List		M
Side relative to tumour	1. Ipsilateral 2. Contralateral 3. Bilateral	Drop-Down List		M
Site	1. IA 2. IB 3. IIA	Drop-Down List *		M

Data	Value	Format	Notes	M
	4. IIB 5. III 6. IV 7. VA 8. VB 9. VI			
Lymphnode development Jugulodigastric area: 3 month growing rate (in %)	quantitative	000,00		
Lymph node development Other cervical areas: 3 month growing rate (in %)	quantitative	000,00		
Selected images	up to 10	jpeg	A number og images selected by the radiologist (apprx 10 images)	

Item 5a Radiomics data from CT scans

Automatically generated by OncoRadiomics SW

Data	Value	Format	Notes	M
GLCM_autocorr	Quantitative	0.00	Gray Level Cooccurance Matrix based features	M
GLCM_clusProm	Quantitative	0.00		M
GLCM_clusShade	Quantitative	0.00		M
GLCM_clusTend	Quantitative	0.00		M
GLCM_contrast	Quantitative	0.00		M
GLCM_correl1	Quantitative	0.00		M
GLCM_diffEntro	Quantitative	0.00		M
GLCM_dissimilar	Quantitative	0.00		M
GLCM_energy	Quantitative	0.00		M
GLCM_entrop2	Quantitative	0.00		M
GLCM_homogeneity1	Quantitative	0.00		M
GLCM_homogeneity2	Quantitative	0.00		M
GLCM_infoCorr1	Quantitative	0.00		M
GLCM_infoCorr2	Quantitative	0.00		M
GLCM_invDiffmomnor	Quantitative	0.00		M
GLCM_invDiffnorm	Quantitative	0.00		M
GLCM_inverseVar	Quantitative	0.00		M
GLCM_maxProb	Quantitative	0.00		M
GLCM_sumAvg	Quantitative	0.00		M

Data	Value	Format	Notes	M
GLCM_sumEntro	Quantitative	0.00		M
GLCM_sumSquares	Quantitative	0.00		M
GLCM_sumVar	Quantitative	0.00		M
GLSZM_highIntensityEmphasis	Quantitative	0.00	Gray Level Size Zone Matrix based features	M
GLSZM_highIntensityLarteAreaEmp	Quantitative	0.00		M
GLSZM_highIntensitySmallAreaEmp	Quantitative	0.00		M
GLSZM_intensityVariability	Quantitative	0.00		M
GLSZM_largeAreaEmphasis	Quantitative	0.00		M
GLSZM_lowIntensityEmphasis	Quantitative	0.00		M
GLSZM_lowIntensityLargeAreaEmp	Quantitative	0.00		M
GLSZM_lowIntensitySmallAreaEmp	Quantitative	0.00		M
GLSZM_sizeZoneVariability	Quantitative	0.00		M
GLSZM_smallAreaEmphasis	Quantitative	0.00		M
GLSZM_zonePercentage	Quantitative	0.00		M
LoG_sigma_XX_mm_3D_stats_energy	Quantitative	0.00	Laplacian of Gaussian Features, XX is the width of LoG filter in mm	M
LoG_sigma_XX_mm_3D_stats_energy_pos	Quantitative	0.00		M
LoG_sigma_XX_mm_3D_stats_entropy	Quantitative	0.00		M
LoG_sigma_XX_mm_3D_stats_entropy_pos	Quantitative	0.00		M
LoG_sigma_XX_mm_3D_stats_kurtosis	Quantitative	0.00		M
LoG_sigma_XX_mm_3D_stats_kurtosis_pos	Quantitative	0.00		M
LoG_sigma_XX_mm_3D_stats_max	Quantitative	0.00		M
LoG_sigma_XX_mm_3D_stats_md	Quantitative	0.00		M
LoG_sigma_XX_mm_3D_stats_md_pos	Quantitative	0.00		M
LoG_sigma_XX_mm_3D_stats_mean	Quantitative	0.00		M
LoG_sigma_XX_mm_3D_stats_mean_pos	Quantitative	0.00		M
LoG_sigma_XX_mm_3D_stats_median	Quantitative	0.00		M
LoG_sigma_XX_mm_3D_stats_median_pos	Quantitative	0.00		M
LoG_sigma_XX_mm_3D_stats_min	Quantitative	0.00		M
LoG_sigma_XX_mm_3D_stats_min_pos	Quantitative	0.00		M
LoG_sigma_XX_mm_3D_stats_range	Quantitative	0.00		M
LoG_sigma_XX_mm_3D_stats_range_pos	Quantitative	0.00		M
LoG_sigma_XX_mm_3D_stats_rms	Quantitative	0.00		M

Data	Value	Format	Notes	M
LoG_sigma_XX_mm_3D_stats_rms_pos	Quantitative	0.00		M
LoG_sigma_XX_mm_3D_stats_skewness	Quantitative	0.00		M
LoG_sigma_XX_mm_3D_stats_skewness_pos	Quantitative	0.00		M
LoG_sigma_XX_mm_3D_stats_std	Quantitative	0.00		M
LoG_sigma_XX_mm_3D_stats_std_pos	Quantitative	0.00		M
LoG_sigma_XX_mm_3D_stats_totalenergy	Quantitative	0.00		M
LoG_sigma_XX_mm_3D_stats_totalenergy_pos	Quantitative	0.00		M
LoG_sigma_XX_mm_3D_stats_uniformity	Quantitative	0.00		M
LoG_sigma_XX_mm_3D_stats_uniformity_pos	Quantitative	0.00		M
LoG_sigma_XX_mm_3D_stats_var	Quantitative	0.00		M
LoG_sigma_XX_mm_3D_stats_var_pos	Quantitative	0.00		M
RLGL_grayLevelNonuniformity	Quantitative	0.00	Run Length Gray Level Matrix based features	M
RLGL_highGrayLevelRunEmphasis	Quantitative	0.00		M
RLGL_longRunEmphasis	Quantitative	0.00		M
RLGL_longRunHighGrayLevEmpha	Quantitative	0.00		M
RLGL_longRunLowGrayLevEmpha	Quantitative	0.00		M
RLGL_lowGrayLevelRunEmphasis	Quantitative	0.00		M
RLGL_runLengthNonuniformity	Quantitative	0.00		M
RLGL_runPercentage	Quantitative	0.00		M
RLGL_shortRunEmphasis	Quantitative	0.00		M
RLGL_shortRunHighGrayLevEmpha	Quantitative	0.00		M
RLGL_shortRunLowGrayLevEmpha	Quantitative	0.00		M
Shape_compactness	Quantitative	0.00	Shape Features	M
Shape_compactness2	Quantitative	0.00		M
Shape_maxDiameter2D1	Quantitative	0.00		M
Shape_maxDiameter2D2	Quantitative	0.00		M
Shape_maxDiameter2D3	Quantitative	0.00		M
Shape_maxDiameter3D	Quantitative	0.00		M
Shape_spherDisprop	Quantitative	0.00		M
Shape_sphericity	Quantitative	0.00		M
Shape_surfVolRatio	Quantitative	0.00		M
Shape_surface	Quantitative	0.00		M
Shape_volume	Quantitative	0.00		M
Shape_volumeNumber	Quantitative	0.00		M
Stats_energy	Quantitative	0.00	First order Features	M
Stats_entropy	Quantitative	0.00		M
Stats_kurtosis	Quantitative	0.00		M

Data	Value	Format	Notes	M
Stats_max	Quantitative	0.00		M
Stats_md	Quantitative	0.00		M
Stats_mean	Quantitative	0.00		M
Stats_median	Quantitative	0.00		M
Stats_min	Quantitative	0.00		M
Stats_peak	Quantitative	0.00		M
Stats_range	Quantitative	0.00		M
Stats_rms	Quantitative	0.00		M
Stats_skewness	Quantitative	0.00		M
Stats_std	Quantitative	0.00		M
Stats_totalenergy	Quantitative	0.00		M
Stats_uniformity	Quantitative	0.00		M
Stats_var	Quantitative	0.00		M
Wavelet_Filter_glcm_autocorr	Quantitative	0.00	Wavelet Based Features ,Filter can be HHH, LLL,HHL, HLL,LHL,LLH,LHH,HLH	M
Wavelet_Filter_glcm_clusProm	Quantitative	0.00		M
Wavelet_Filter_glcm_clusShade	Quantitative	0.00		M
Wavelet_Filter_glcm_clusTend	Quantitative	0.00		M
Wavelet_Filter_glcm_contrast	Quantitative	0.00		M
Wavelet_Filter_glcm_correl1	Quantitative	0.00		M
Wavelet_Filter_glcm_diffEntro	Quantitative	0.00		M
Wavelet_Filter_glcm_dissimilar	Quantitative	0.00		M
Wavelet_Filter_glcm_energy	Quantitative	0.00		M
Wavelet_Filter_glcm_entrop2	Quantitative	0.00		M
Wavelet_Filter_glcm_homogeneity1	Quantitative	0.00		M
Wavelet_Filter_glcm_homogeneity2	Quantitative	0.00		M
Wavelet_Filter_glcm_infoCorr1	Quantitative	0.00		M
Wavelet_Filter_glcm_infoCorr2	Quantitative	0.00		M
Wavelet_Filter_glcm_invDiffmomnor	Quantitative	0.00		M
Wavelet_Filter_glcm_invDiffnorm	Quantitative	0.00		M
Wavelet_Filter_glcm_inverseVar	Quantitative	0.00		M
Wavelet_Filter_glcm_maxProb	Quantitative	0.00		M
Wavelet_Filter_glcm_sumAvg	Quantitative	0.00		M
Wavelet_Filter_glcm_sumEntro	Quantitative	0.00		M
Wavelet_Filter_glcm_sumSquares	Quantitative	0.00		M
Wavelet_Filter_glcm_sumVar	Quantitative	0.00		M
Wavelet_Filter_glszm_highIntensityEmphasis	Quantitative	0.00		M
Wavelet_Filter_glszm_highIntensityLarteAreaEmp	Quantitative	0.00		M
Wavelet_Filter_glszm_highIntensitySmallAreaEmp	Quantitative	0.00		M
Wavelet_Filter_glszm_intensityVaria	Quantitative	0.00		M

Data	Value	Format	Notes	M
bility				
Wavelet_Filter_glszm_largeAreaEmphasis	Quantitative	0.00		M
Wavelet_Filter_glszm_lowIntensityEmphasis	Quantitative	0.00		M
Wavelet_Filter_glszm_lowIntensityLargeAreaEmp	Quantitative	0.00		M
Wavelet_Filter_glszm_lowIntensitySmallAreaEmp	Quantitative	0.00		M
Wavelet_Filter_glszm_sizeZoneVariability	Quantitative	0.00		M
Wavelet_Filter_glszm_smallAreaEmphasis	Quantitative	0.00		M
Wavelet_Filter_glszm_zonePercentage	Quantitative	0.00		M
Wavelet_Filter_rlgl_grayLevelNonuniformity	Quantitative	0.00		M
Wavelet_Filter_rlgl_highGrayLevelRunEmphasis	Quantitative	0.00		M
Wavelet_Filter_rlgl_longRunEmphasis	Quantitative	0.00		M
Wavelet_Filter_rlgl_longRunHighGrayLevEmpha	Quantitative	0.00		M
Wavelet_Filter_rlgl_longRunLowGrayLevEmpha	Quantitative	0.00		M
Wavelet_Filter_rlgl_lowGrayLevelRunEmphasis	Quantitative	0.00		M
Wavelet_Filter_rlgl_runLengthNonuniformity	Quantitative	0.00		M
Wavelet_Filter_rlgl_runPercentage	Quantitative	0.00		M
Wavelet_Filter_rlgl_shortRunEmphasis	Quantitative	0.00		M
Wavelet_Filter_rlgl_shortRunHighGrayLevEmpha	Quantitative	0.00		M
Wavelet_Filter_rlgl_shortRunLowGrayLevEmpha	Quantitative	0.00		M
Wavelet_Filter_stats_energy	Quantitative	0.00		M
Wavelet_Filter_stats_entropy	Quantitative	0.00		M
Wavelet_Filter_stats_kurtosis	Quantitative	0.00		M
Wavelet_Filter_stats_max	Quantitative	0.00		M
Wavelet_Filter_stats_md	Quantitative	0.00		M
Wavelet_Filter_stats_mean	Quantitative	0.00		M
Wavelet_Filter_stats_median	Quantitative	0.00		M
Wavelet_Filter_stats_min	Quantitative	0.00		M
Wavelet_Filter_stats_range	Quantitative	0.00		M
Wavelet_Filter_stats_rms	Quantitative	0.00		M
Wavelet_Filter_stats_skewness	Quantitative	0.00		M
Wavelet_Filter_stats_std	Quantitative	0.00		M
Wavelet_Filter_stats_totalenergy	Quantitative	0.00		M

Data	Value	Format	Notes	M
Wavelet_Filter_stats_uniformity	Quantitative	0.00		M
Wavelet_Filter_stats_var	Quantitative	0.00		M
Intensity Volume Histogram based Features (#233)	Quantitative	0.00	PET only	M

NOTE:

For Research purpose, Oncoradiomics software can provide 400+ Radiomics features that are then associated to the patients in the Big data repository and can be accessed by the researchers. A selection of features which have prognostic value for head and neck cancer have to be isolated by the Big Data Analysis and integrated in the clinical decision support system.

The Big Data Analysis will extract the relevant informative features to be used by the models.

Item 5b Radiomics data from MRI and DWI-MRI scans

Automatically generated by the radiomics SW developed by POLIMI

Data	Value	Format	Notes	M
S&S_Compactness 1	quantitative	000,00;	feature value	M
S&S_Compactness 2	quantitative	000,00;	feature value	M
S&S_Maximum 3D diameter	quantitative	000,00;	feature value	M
S&S_Minimum 3D diameter	quantitative	000,00;	feature value	M
S&S_Median 3D diameter	quantitative	000,00;	feature value	M
S&S_Mean 3D diameter	quantitative	000,00;	feature value	M
S&S_Std 3D diameter	quantitative	000,00;	feature value	M
S&S_Variance 3D diameter	quantitative	000,00;	feature value	M
S&S_Skewness 3D diameter	quantitative	000,00;	feature value	M
S&S_Kurtosis 3D diameter	quantitative	000,00;	feature value	M
S&S_Equivalent R	quantitative	000,00;	feature value	M
S&S_Max (D/2) / R	quantitative	000,00;	feature value	M
S&S_Spherical disproportion	quantitative	000,00;	feature value	M
S&S_Sphericity	quantitative	000,00;	feature value	M
S&S_Area	quantitative	000,00;	feature value	M
S&S_Surface to volume ratio	quantitative	000,00;	feature value	M
S&S_Volume	quantitative	000,00;	feature value	M
S&S_mean facet normal0	quantitative	000,00;	feature value	M
S&S_mean facet normal1	quantitative	000,00;	feature value	M
S&S_mean facet normal2	quantitative	000,00;	feature value	M
FOS_Signal Energy	quantitative	000,00;	feature value	M
FOS_Signal Kurtosis	quantitative	000,00;	feature value	M
FOS_Signal Mad	quantitative	000,00;	feature value	M

Data	Value	Format	Notes	M
FOS_Signal Max	quantitative	000,00;	feature value	M
FOS_Signal Mean	quantitative	000,00;	feature value	M
FOS_Signal Median	quantitative	000,00;	feature value	M
FOS_Signal Min	quantitative	000,00;	feature value	M
FOS_Signal Range	quantitative	000,00;	feature value	M
FOS_Signal RMS	quantitative	000,00;	feature value	M
FOS_Signal Skewness	quantitative	000,00;	feature value	M
FOS_Signal STD	quantitative	000,00;	feature value	M
FOS_Signal Variance	quantitative	000,00;	feature value	M
FOS_Histogram Entropy	quantitative	000,00;	feature value	M
FOS_Histogram Kurtosis	quantitative	000,00;	feature value	M
FOS_Histogram Mad	quantitative	000,00;	feature value	M
FOS_Histogram Max	quantitative	000,00;	feature value	M
FOS_Histogram Mean	quantitative	000,00;	feature value	M
FOS_Histogram Median	quantitative	000,00;	feature value	M
FOS_Histogram Min	quantitative	000,00;	feature value	M
FOS_Histogram Range	quantitative	000,00;	feature value	M
FOS_Histogram RMS	quantitative	000,00;	feature value	M
FOS_Histogram Skewness	quantitative	000,00;	feature value	M
FOS_Histogram STD	quantitative	000,00;	feature value	M
FOS_Histogram Uniformity	quantitative	000,00;	feature value	M
FOS_Histogram Variance	quantitative	000,00;	feature value	M
FOS_Histogram TotalFrequency	quantitative	000,00;	feature value	M
FOS_Histogram Quantile 0	quantitative	000,00;	feature value	M
FOS_Histogram Quantile 0.1	quantitative	000,00;	feature value	M
FOS_Histogram Quantile 0.2	quantitative	000,00;	feature value	M
FOS_Histogram Quantile 0.3	quantitative	000,00;	feature value	M
FOS_Histogram Quantile 0.4	quantitative	000,00;	feature value	M
FOS_Histogram Quantile 0.5	quantitative	000,00;	feature value	M
FOS_Histogram Quantile 0.6	quantitative	000,00;	feature value	M
FOS_Histogram Quantile 0.7	quantitative	000,00;	feature value	M
FOS_Histogram Quantile 0.8	quantitative	000,00;	feature value	M
FOS_Histogram Quantile 0.9	quantitative	000,00;	feature value	M
FOS_Histogram Quantile 1	quantitative	000,00;	feature value	M
GLCM_Autocorrelation	quantitative	000,00;	feature value	M
GLCM_Cluster Prominence	quantitative	000,00;	feature value	M
GLCM_Cluster Shade	quantitative	000,00;	feature value	M
GLCM_Cluster Tendency	quantitative	000,00;	feature value	M
GLCM_Contrast	quantitative	000,00;	feature value	M
GLCM_Correlation	quantitative	000,00;	feature value	M
GLCM_Difference Entropy	quantitative	000,00;	feature value	M
GLCM_Dissimilarity	quantitative	000,00;	feature value	M
GLCM_Energy	quantitative	000,00;	feature value	M
GLCM_Entropy	quantitative	000,00;	feature value	M

Data	Value	Format	Notes	M
GLCM_Homogeneity	quantitative	000,00;	feature value	M
GLCM_homogeneity2	quantitative	000,00;	feature value	M
GLCM_IMOC1	quantitative	000,00;	feature value	M
GLCM_IMOC2	quantitative	000,00;	feature value	M
GLCM_Inverse Difference moment	quantitative	000,00;	feature value	M
GLCM_Inverse Difference moment2	quantitative	000,00;	feature value	M
GLCM_Inverse Variance	quantitative	000,00;	feature value	M
GLCM_Max Probability	quantitative	000,00;	feature value	M
GLCM_Sum Average	quantitative	000,00;	feature value	M
GLCM_Sum Entropy	quantitative	000,00;	feature value	M
GLCM_Inertia	quantitative	000,00;	feature value	M
GLRLM_Short Run Emphasis	quantitative	000,00;	feature value	M
GLRLM_Long Run Emphasis	quantitative	000,00;	feature value	M
GLRLM_Gray Level Non Uniformity	quantitative	000,00;	feature value	M
GLRLM_Run Length Non Uniformity	quantitative	000,00;	feature value	M
GLRLM_Run Percentage	quantitative	000,00;	feature value	M
GLRLM_Low Gray Level Run Emphasis	quantitative	000,00;	feature value	M
GLRLM_high Gray Level Run Emphasis	quantitative	000,00;	feature value	M
GLRLM_Short Run Low Gray Level Emphasis	quantitative	000,00;	feature value	M
GLRLM_Short Run High Gray Level Emphasis	quantitative	000,00;	feature value	M
GLRLM_Long Run Low Gray Level Emphasis	quantitative	000,00;	feature value	M
GLRLM_Long Run High Gray Level Emphasis	quantitative	000,00;	feature value	M
WLTHHH_* (69 features)	quantitative	000,00;	1. WLTHHH_FOS_Signal_Energy 2. WLTHHH_FOS_Signal_Kurtosis 3. …….. ….. 68. …… 69. WLTHH_GLRLM_Long Run High Gray Level Emphasis	M
WLTHHL_* (69 features)	quantitative	000,00;	1. WLTHHL_FOS_Signal_E	M

Data	Value	Format	Notes	M
			nergy 2. WLTHHL_FOS_Signal_Kurtosis 3. …….. ….. 68. …… 69. WLTHHL_GLRLM_Long Run High Gray Level Emphasis	
WLTHLH_* (69 features)	quantitative	000,00;	1. WLTHLH_FOS_Signal_Energy 2. WLTHLH_FOS_Signal_Kurtosis 3. …….. ….. 68. …… 69. WLTLH_GLRLM_Long Run High Gray Level Emphasis	M
WLTHLL_* (69 features)	Quantitative	000,00;	1. WLTHLL_FOS_Signal_Energy 2. WLTHLL_FOS_Signal_Kurtosis 3. …….. ….. 68. …… 69. WLTLL_GLRLM_Long Run High Gray Level Emphasis	M
WLTLHH_* (69 features)	quantitative	000,00;	1. WLTLHH_FOS_Signal_Energy 2. WLTLHH_FOS_Signal_Kurtosis 3. …….. ….. 68. …… 69. WLTLHH_GLRLM_Long Run High Gray Level Emphasis	M

Data	Value	Format	Notes	M
WLTLHL_* (69 features)	quantitative	000,00;	1. WLTLHL_FOS_Signal_Energy 2. WLTLHL_FOS_Signal_Kurtosis 3. …….. ….. 68. …… 69. WLTLHL_GLRLM_Long Run High Gray Level Emphasis	M
WLTLLH_* (69 features)	quantitative	000,00;	1. WLTLLH_FOS_Signal_Energy 2. WLTLLH_FOS_Signal_Kurtosis 3. …….. ….. 68. …… 69. WLTLLH_GLRLM_Long Run High Gray Level Emphasis	M
WLTLLL_* (69 features)	quantitative	000,00;	1. WLTLLL_FOS_Signal_Energy 2. WLTLLL_FOS_Signal_Kurtosis 3. …….. ….. 68. …… 69. WLTLLL_GLRLM_Long Run High Gray Level Emphasis	M

NOTE: The Big Data Analysis will extract the relevant informative features to be used by the models.

Item 6 - Pathology data

Data	Value	Format	Notes	M
Pathology case number	text	Text-box		M
Date of analysis	date	DD/MM/YYYY		M
Previous relevant pathological samples	text	Text-box	# reference biopsy	
Site	ICD10 definition	Select List	Ref. Addendum A	M
Tumor maximum diameter (mm)	quantitative	000,00		M
Tumor thickness (mm)	quantitative	000,00		M
Depth of invasion (mm)	quantitative	000,00		M
Lymphocytic infiltration If oropharynx	1. No 2. Yes	Radio-Button		
Level of invasion	1. Microinvasion 2. Lamina propria 3. Submucosa 4. Muscle	Drop-Down List *		
Pattern of invasion	1. Cohesive 2. Focal non cohesive <50% 3. Extensive non cohesive	Drop-Down List *		
Basaloid features	1. No 2. Yes	Radio-Button		M
Lympho-plasmacytic reaction	1. None 2. Mild 3. Moderate 4. Marked	Drop-Down List		
Lympho-vascular invasion	1. No 2. Yes	Radio-Button		M
Perineural invasion	1. No 2. Focal (1 slides) 3. Extensive (all slides)	Drop-Down List		M
Degree of cell keratinisation	1. None 2. Mild 3. Moderate 4. High	Drop-Down List		
Nuclear pleomorphism	1. None 2. Mild 3. Moderate 4. Marked 5. Extreme	Drop-Down List		
Number of mitoses per 10 High-power fields (HPFs)	quantitative	000		
Grade at diagnosis:	1. Gx 2. G1 low grade	Drop-Down list *	Gx: undetermined G1: well	M

Data	Value	Format	Notes	M
	3. G2 intermediate grade 4. G3 high grade 5. G4 high grade 6. Not Done 7. Unknown 8. Not Applicable		differentiated G2: moderately differentiated G3: poorly differentiated G4: undifferentiated	
TNM pT	1. Tis 2. T1 3. T2 4. T3 5. T4a 6. T4b 7. T4c	Drop-Down list *		M
TNM pN	1. N0 2. N1 3. N2a 4. N2b 5. N2c 6. N3	Drop-Down list *		M
Number of metastatic lymphnodes right side level I	quantitative	00	IF NOT COMPLETED =0	
Number of metastatic lymphnodes right side level II	quantitative	00	IF NOT COMPLETED =0	
Number of metastatic lymphnodes right side level III	quantitative	00	IF NOT COMPLETED =0	
Number of metastatic lymphnodes right side level IV	quantitative	00	IF NOT COMPLETED =0	
Number of metastatic lymphnodes right side level V	quantitative	00	IF NOT COMPLETED =0	
Number of metastatic lymphnodes left side level I	quantitative	00	IF NOT COMPLETED =0	
Number of metastatic lymphnodes left side level II	quantitative	00	IF NOT COMPLETED =0	
Number of metastatic lymphnodes left side level III	quantitative	00	IF NOT COMPLETED =0	
Number of metastatic lymphnodes left side level IV	quantitative	00	IF NOT COMPLETED =0	
Number of metastatic lymphnodes left side level V	quantitative	00	IF NOT COMPLETED =0	
Extracapsular spread	1. No 2. Expansive 3. Infiltrative 4. Not Applicable	Drop-Down list	Default = No	M
Surgical margins	1. Involved (<1 mm) 2. Close (1-5 mm)	Drop-Down list		M

Data	Value	Format	Notes	M
	3. Clear (>5 mm)			
p53 stain	2, 3, 4, 5, 6, 7, 8, 9	Drop-Down list		
p16ink4a stain >70% tumor (indirect HPV marker)	1. Negative 2. Positive	Drop-Down list		
Ki67 stain	2, 3, 4, 5, 6, 7, 8, 9	Drop-Down list		
Pan Cytokeratin	1. Negative 2. Positive	Drop-Down list		
HPV DNA	1. Negative 2. Positive	Drop-Down list		M

Item 7 - Chemotherapy

Data	Value	Format	Notes
Cancer Therapy Agent Name	1. CDDP 2. Carboplatin 3. Cetuximab 4. Cisplatin 5. Docetaxel 6. Methotrexate 7. Paclitaxel 8. Other 9. Unknown	Select List *	
Type of CT	1. Induction 2. Concomitant 3. Adjuvant	Drop-Down List *	
Starting Dose (mg/m2)	quantitative	00	
Treatment Scheduled	1. Daily 2. Weekly 3. Every 21 days 4. Every 28 days 5. Other	*	
Number of cycles performed	quantitative	00	Value/Format: "Numeric"
Start Date of First Cycle	date	DD/MM/YYYY	
Stop Date of Last Cycle	date	DD/MM/YYYY	
What was the reason for stopping the treatment?	1. Completed Treatment 2. Treatment complication/Adverse Event/Adverse Drug Reaction 3. Progression 4. Physician decision 5. Death 6. Other 7. Not specified	Select List *	
Best Tumor Response	1. Complete response 2. Partial response 3. Stable disease 4. Progression 5. Non-evaluable	Drop-Down List *	
Note	text	Text-box	

Item 8 – Radiotherapy

Data	Value	Format	Notes
Radiotherapy Area	text	Text-box	
Type of RT	1. 3D-RT 2. IMRT 3. Other	**DROP DOWN LIST ***	
Note to type of RT	text	Text-box	
Setting	1. Definitive 2. Preoperative 3. Postoperative 4. Salvage	Drop-Down List	
RT start date	date	DD/MM/YYYY	
RT end date	date	DD/MM/YYYY	
Overall treatment Time OTT (months)	quantitative	000	Automatic calculation: (RT end date) - (RT start date)
TPT (for postoperative cases) (months)	quantitative	000	Automatic calculation: (Date of Surgery) - (RT start date)
Dose to HR-PTV (Gy)	1. 64 2. 66 3. 68 4. 70 5. Other	Drop-Down List *	
Dose for IR-PTV (Gy):	1. 60 2. 64 3. Other	Drop-Down List *	
Dose to LR-PTV (Gy)	1. 45 2. 50 3. Other	Drop-Down List *	
Area for N included in HR-PTV	1. Monolateral neck 2. Bilateral neck	Drop-Down List	
Fractionation	1. Conventional (up to 2 Gy/fr) 2. Moderately hypofractionated (up to 2.2 Gy/fr 3. Accelerated 4. Hyperfractionated 5. Other	Drop-Down List *	
Reason for discontinuation	1. Treatment completed 2. Progression 3. Toxicity 4. Other	Drop-Down List	
Note	text	Text-box	

Item 9 - Surgery

Data	Value	Format	Notes	M
Has the patient undergone any cancer surgery for SCCHN since diagnosis?	1. No 2. Yes	Radio-Button	THIS FIELD MAY BE OMITTED AND REPLACED BY THE DATE OF SURGERY	M
NOTE: THE FOLLOWING DATA ARE COMPLETED ONLY IF SURGERY IS YES				
Date of Surgery	date	DD/MM/YYYY	IF BLANK= NO SURGERY	M
Type of surgery	1. Laryngectomy 2. Pharynectomy 3. Laryngopharyngectomy 4. Neck dissection 5. (partial) Glossectomy 6. Resection 7. Other	Select list *	Dependent if surgery=yes / date of surgery not blank	M
Surgery site	1. Primary Site 2. Lymph Node 3. Distant Metastatic 4. Site (specify)	Select list		M
Note to surgery site	text	Text-box	DEPENDING IF SURGERY SITE=4	
Is this surgery related to primary tumor or recurrence or other?	1. Primary Tumor 2. Recurrence 3. Other	Drop-Down List		M
Flap reconstruction	1. No 2. Yes	Radio-Button	DEFAULT = No	
Reconstruction	1. Primary closure 2. Local flaps 3. Loco-regional flaps 4. Distant pedicled flaps 5. Free flap 6. Graft	Select list *		
Hospitalization start date	date	DD/MM/YYYY		M
Hospitalization end date	date	DD/MM/YYYY		M
Complication	1. No 2. Yes	Radio-Button		M
If yes, specify what complications	1. Wound infection 2. Intraoperative hemorrhage	Select List *	Dependent from choice 2. in “Complication”	

Data	Value	Format	Notes	M
	3. Postoperative hemorrhage 4. Wound complication 5. Wound dehiscence 6. Skin loss 7. Flap necrosis 8. Fistula 9. Other			
What action taken	1. Further surgery 2. Nothing waiting and observation 3. Hyperbaric radiotherapy 4. Other	Select List *	Dependent from choice 2. in "Complication" If "yes" in "complications"	
Date of resolution	date	DD/MM/YYYY	Dependent from choice 2. in "Complication"	
Did complication modify plan of care?	1. No 2. Yes	Radio-Button	Dependent from choice 2. in "Complication"	
Note to how complication modified plan of care	text	Text-box	Dependent from choice 2. in "Did complication modify plan of care"	
Neck dissection	1. No 2. Yes	Radio-Button		M
Site neck dissection	1. Right 2. Left 3. Bilateral	Drop-Down List	Dependent from choice 2. in "Neck dissection"	
Kind of neck dissection	4. Selective 5. Functional 6. Radical modified 7. Radical	Drop-Down List	Dependent from choice 2. in "Neck dissection"	

Item 10 - Tissue Sample

Data	Value	Format	Notes	M
Date tumor sample was taken:	date	DD/MM/YYYY		M
Type of tumor sample	1. Surgical resection 2. Surgical biopsy 3. Punch biopsy 4. Other	Drop-Down List *		M
Tumor Location	1. Primary Tumor 2. Tumor tissue from recurrent disease 3. Tumor tissue from metastatic disease	Drop-Down List		M
Site of tumor sample	1. Oral Cavity 2. Oropharnynx 3. Hypopharynx 4. Larynx 5. Lymph node (specify) 6. Distant metastatic site (specify)	Select List *		M

Item 10a - Genomic data

Automatically collected by the genomic analysis. The detailed data shall be available for Big Data Analysis

Data	Value	Format	Notes	M/O
ID Analysis	Text (ID Sample)	text-area	Unique Sample ID	M
Date of Analysis (RNA extraction)	Text	date		O
Operator Name (RNA extraction)	Text	text-area		O
Date of Analysis (library construction and sequencing)	Text	date		O
Operator Name (library construction and sequencing)	Text	text-area		O
Date of Analysis (data processing)	Text	date		M
Operator Name (data processing)	Text	text-area		M
Raw data (entire gene profiling)	BAM file	binary format for storing sequence data	Data storing (please be aware about the size)	M
Raw data (entire gene profiling)	FASTQ file	text-based format for storing both a biological sequence and its corresponding quality scores	Data storing (please note be aware about the size)	M
Processed data (entire gene profiling)	tab-delimited files	quantitative	Data storing	M
Reference Gene (raw count)	number	quantitative	Data storing	M
Reference Gene (normalized count)	number	quantitative	Data storing	M

Genes relevant for prognosis

Data	Value	Format	Notes	M/O
Target Gene Name	Text	text-area	Data matrix for analysis	M
Gene annotation (Gene Symbol)	Text	text-area	Data matrix for analysis	M
Gene annotation (EntrezID)	Text	text-area	Data matrix for analysis	M
Reference Gene (raw count)	number	quantitative	Data matrix for analysis	M
Reference Gene (normalized count)	number	quantitative	Data matrix for analysis	M

NOTE: Only genes relevant for prognosis. The rows in green are repeated for each prognostic gene used by the models. O: optional; M: mandatory

Item 11 - Follow up

Data	Value	Format	Notes	M
Date of examination	date	DD/MM/YYYY		M
Follow-up period (months)	quantitative	000	Automatic calculation: (DoD) - (date of examination)	A
Status of Patient	1. Dead 2. Alive	Radio-Button		M
Date of death	date	DD/MM/YYYY	Dependent from choice 1. in "Status of Patient"	
Primary cause of death	text	Text-box	Dependent from choice 1. in "Status of Patient"	
Date of last contact:	date	DD/MM/YYYY	Default = date of examination	
Recurrence	1. No 2. Yes	Radio-Button		M
Type recurrence	1. Local 2. Regional 3. Second primary tumor 4. Distant metastasis 5. Delayed lymp node metastatis	Drop-Down List	Dependent from choice 2. in "Recurrence"	
Follow-up period before recurrence	date	DD/MM/YYYY	Automatic calculation: (DoD) - (date of examination); dependent from choice 2. in "Recurrence"	
Treatment of recurrence	1. Surgery 2. RT 3. CRT 4. CT 5. Clinical study 6. Other	Select List	Dependent from choice 2. in "Recurrence"	
Intention	1. Curative 2. Palliative	Drop-Down List	Dependent from choice 2. in "Recurrence"	
Quality of life at Last Evaluation				
Deglutition and respiration	1. Normal 2. Liquid diet 3. Semi-liquid diet 4. PEG 5. SNG 6. Permanent tracheotomy	Drop-Down List		M

Item 12 – Toxicity

(according to CTCAE V4) - NOTE: this item is present only if radio or chemotherapy

Acute Toxicity since treatment (CT/CTRT) start to 30 days after treatment end				
Data	**Value**	**Format**	**Notes**	**M**
Infection ≥ G3	1. No 2. Yes	Radio-Button	Default = No	M
Mucositis ≥ G3	1. No 2. Yes	Radio-Button	Default = No	M
Medullary toxicity ≥ G3	1. No 2. Yes	Radio-Button	Default = No	M
Hepatic toxicity ≥ G3	1. No 2. Yes	Radio-Button	Default = No	M
Renal toxicity ≥ G3	1. No 2. Yes	Radio-Button	Default = No	M
Xerostomia ≥ G3	1. No 2. Yes	Radio-Button	Default = No	M
Dysphagia ≥ G3	1. No 2. Yes	Radio-Button	Default = No	M
Late Toxicity at 6, 12 and 24 months till 2 years after treatment end				
Data	**Value**	**Format**	**Notes**	**M**
Disphagia ≥ G3	1. No 2. Yes	Radio-Button	Default = No	M
Dyspnea ≥ G3	1. No 2. Yes	Radio-Button	Default = No	M
Xerostomia ≥ G3	1. No 2. Yes	Radio-Button	Default = No	M
Aspiration ≥ G3	1. No 2. Yes	Radio-Button	Default = No	M
Trismus ≥ G3	1. No 2. Yes	Radio-Button	Default = No	M
Tracheotomy	1. No 2. Yes	Radio-Button	Default = No	M
Gastrostomy / Feeding tube	1. No 2. Yes	Radio-Button	Default = No	M
Trismus ≥ G3	3. No 4. Yes	Radio-Button	Default = No	M
Asthenia ≥ G3	5. No 6. Yes	Radio-Button	Default = No	M
Pain ≥ G3	7. No 8. Yes	Radio-Button	Default = No	M

Item 13 – Quality of Life questionnaires

The questionnaires are the following.

EORTC QLQ-H&N30

EORTC QLQ-H&N30

EuroQol EQ-5D

ADDENDUM A: 2016 ICD-10-CM Diagnosis Codes (http://www.icd10data.com/ICD10CM/Codes)

- C00 Malignant neoplasm of lip
  - C00.0 Malignant neoplasm of external upper lip
  - C00.1 Malignant neoplasm of external lower lip
  - C00.2 Malignant neoplasm of external lip, unspecified
  - C00.3 Malignant neoplasm of upper lip, inner aspect
  - C00.4 Malignant neoplasm of lower lip, inner aspect
  - C00.5 Malignant neoplasm of lip, unspecified, inner aspect
  - C00.6 Malignant neoplasm of commissure of lip, unspecified
  - C00.8 Malignant neoplasm of overlapping sites of lip
  - C00.9 Malignant neoplasm of lip, unspecified
- C01 Malignant neoplasm of base of tongue
- C02 Malignant neoplasm of other and unspecified parts of tongue
  - C02.0 Malignant neoplasm of dorsal surface of tongue
  - C02.1 Malignant neoplasm of border of tongue
  - C02.2 Malignant neoplasm of ventral surface of tongue
  - C02.3 Malignant neoplasm of anterior two-thirds of tongue, part unspecified
  - C02.4 Malignant neoplasm of lingual tonsil
  - C02.8 Malignant neoplasm of overlapping sites of tongue
  - C02.9 Malignant neoplasm of tongue, unspecified
- C03 Malignant neoplasm of gum
  - C03.0 Malignant neoplasm of upper gum
  - C03.1 Malignant neoplasm of lower gum

- C03.9 Malignant neoplasm of gum, unspecified
- C04 Malignant neoplasm of floor of mouth
  - C04.0 Malignant neoplasm of anterior floor of mouth
  - C04.1 Malignant neoplasm of lateral floor of mouth
  - C04.8 Malignant neoplasm of overlapping sites of floor of mouth
  - C04.9 Malignant neoplasm of floor of mouth, unspecified
- C05 Malignant neoplasm of palate
  - C05.0 Malignant neoplasm of hard palate
  - C05.1 Malignant neoplasm of soft palate
  - C05.2 Malignant neoplasm of uvula
  - C05.8 Malignant neoplasm of overlapping sites of palate
  - C05.9 Malignant neoplasm of palate, unspecified
- C06 Malignant neoplasm of other and unspecified parts of mouth
  - C06.0 Malignant neoplasm of cheek mucosa
  - C06.1 Malignant neoplasm of vestibule of mouth
  - C06.2 Malignant neoplasm of retromolar area
  - C06.8 Malignant neoplasm of overlapping sites of other and unspecified parts of mouth
    - C06.80 Malignant neoplasm of overlapping sites of unspecified parts of mouth
    - C06.89 Malignant neoplasm of overlapping sites of other parts of mouth
      - C06.9 Malignant neoplasm of mouth, unspecified
- C07 Malignant neoplasm of parotid gland
- C08 Malignant neoplasm of other and unspecified major salivary glands
  - C08.0 Malignant neoplasm of submandibular gland
  - C08.1 Malignant neoplasm of sublingual gland
  - C08.9 Malignant neoplasm of major salivary gland, unspecified
- C09 Malignant neoplasm of tonsil
  - C09.0 Malignant neoplasm of tonsillar fossa
  - C09.1 Malignant neoplasm of tonsillar pillar (anterior) (posterior)
  - C09.8 Malignant neoplasm of overlapping sites of tonsil
  - C09.9 Malignant neoplasm of tonsil, unspecified
- C10 Malignant neoplasm of oropharynx
  - C10.0 Malignant neoplasm of vallecula
  - C10.1 Malignant neoplasm of anterior surface of epiglottis

- C10.2 Malignant neoplasm of lateral wall of oropharynx
- C10.3 Malignant neoplasm of posterior wall of oropharynx
- C10.4 Malignant neoplasm of branchial cleft
- C10.8 Malignant neoplasm of overlapping sites of oropharynx
- C10.9 Malignant neoplasm of oropharynx, unspecified

C11 Malignant neoplasm of nasopharynx

- C11.0 Malignant neoplasm of superior wall of nasopharynx
- C11.1 Malignant neoplasm of posterior wall of nasopharynx
- C11.2 Malignant neoplasm of lateral wall of nasopharynx
- C11.3 Malignant neoplasm of anterior wall of nasopharynx
- C11.8 Malignant neoplasm of overlapping sites of nasopharynx
- C11.9 Malignant neoplasm of nasopharynx, unspecified

C12 Malignant neoplasm of pyriform sinus

C13 Malignant neoplasm of hypopharynx

- C13.0 Malignant neoplasm of postcricoid region
- C13.1 Malignant neoplasm of aryepiglottic fold, hypopharyngeal aspect
- C13.2 Malignant neoplasm of posterior wall of hypopharynx
- C13.8 Malignant neoplasm of overlapping sites of hypopharynx
- C13.9 Malignant neoplasm of hypopharynx, unspecified

C14 Malignant neoplasm of other and ill-defined sites in the lip, oral cavity and pharynx

- C14.0 Malignant neoplasm of pharynx, unspecified
- C14.2 Malignant neoplasm of Waldeyer's ring
- C14.8 Malignant neoplasm of overlapping sites of lip, oral cavity and pharynx

C32 Malignant neoplasm of larynx

- C32.0 Malignant neoplasm of glottis
- C32.1 Malignant neoplasm of supraglottis
- C32.2 Malignant neoplasm of subglottis
- C32.3 Malignant neoplasm of laryngeal cartilage
- C32.8 Malignant neoplasm of overlapping sites of larynx
- C32.9 Malignant neoplasm of larynx, unspecified

ANNEX III - POPULATION DATA

Tumor registries data (Epidemiology data)

Data	Data sources	Provider(s)
• Gender • Date of Birth • Country of residence • Date of diagnose • Primary tumour site and subsite • Morphology code • Smoking (Former smoker, Active smoker, Never smoke) • Comorbidity • HPV test done • Clinical TNM • Pathological TNM • CT scan • MRI scan • Hospital of diagnose • Surgery of primary tumour (date, hospital, resection status) • Post surgery pathology report with the core items available • Chemotherapy (date, hospital) • Radiotherapy (date, hospital) • Follow-up (vital status)	Population-based cancer registries contributing to the RARECAREnet high resolution study	INT, Tumor Registries
• Gender • Date of Birth • Country of residence • Date of diagnosis • Hospital of diagnosis • Topography • Morphology • Stage TNM • Stage Extent of Disease: - Local (confined to the site of origin); - Regional (spread to immediately adjacent tissues and/or regional lymph-nodes); - Metastatic (spread to distant organ) • Treatments (date, type, hospital) • Vital status	Population-based cancer registries contributing to the RARECAREnet pilot study	INT, Tumor Registries
• Gender • Date of Birth • Country of residence • Date of diagnosis • Topography	Population based cancer registries contributing to the EUROCARE (RARECAREnet) survival study	INT

Data	Data sources	Provider(s)
• Morphology • Stage Extent of Disease • Vital status		

Table 1. Data from cancer registries and RARECARENET studies

Lifestyle behaviours[7]

Data	Data sources	Provider(s)
Region - Alcohol consumption (% population) - Alcohol consumption not during meals (% population) - Usual heavy drinkers (% population) - Higher risk drinkers (% population) - Binge consumption (% population) - Higher risk drinkers who have been advised by clinicians to stop	Alcohol consumption. Data from Italy by region. (% of population) Studio PASSI 2011-2014 www.epicentro.iss.it/passi/dati/alcol.asp Population aged 18-69	ISS
Age range - % of alcohol consumers - IC95% inferior - IC95% superior	Alcohol consumption by age, Gender, Level of education, level of income, citizenship. Data from Italy www.epicentro.iss.it/passi/dati/alcol.asp	ISS
Gender (Male/Female) - % of alcohol consumers - IC95% inferior - IC95% superior		
Level of education (No/primary school; secondary school; college; university degree) - % of alcohol consumers - IC95% inferior - IC95% superior		
Economic status (high economic difficulty, some economic difficulty, no economic difficulty) - % of alcohol consumers - IC95% inferior - IC95% superior		

[7] These data are aggregated by geographic area, gender, age range and are based on statistical polls.

Data	Data sources	Provider(s)
Citizenship (Italian, Other Country) - % of alcohol consumers - IC95% inferior - IC95% superior		
Age range - % of alcohol consumers - IC95% inferior - IC95% superior	High risk alcohol consumption (Italy) www.epicentro.iss.it/passi/dati/alcol.asp	ISS
Gender (Male/Female) - % of alcohol consumers - IC95% inferior - IC95% superior		
Level of education (No/primary school; secondary school; college; university degree) - % of alcohol consumers - IC95% inferior - IC95% superior		
Economic status (high economic difficulty, some economic difficulty, no economic difficulty) - % of alcohol consumers - IC95% inferior - IC95% superior		
Citizenship (Italian, Other Country) - % of alcohol consumers - IC95% inferior - IC95% superior		

Table 2. Alcohol consumption by Region (Italy), for each Italian Region and aggregated

Data	Data sources	Provider(s)
Region - Smokers[8] (% population) - Ex smokers (% population) - People who have been advised to	Smoking habits by Region (Italy) www.epicentro.iss.it/passi/dati/fumo.asp	ISS

[8] Smoker: person who has smoked >100 cigarettes in life and still smokes or stopped smoking less than 6 months ago

Data	Data sources	Provider(s)
stop smoking		
Non smokers Smokers - Suspended smoking - Occasional smoker - Daily smoker Ex-smokers Average nb. cigarettes smoked/year	Smoking habits Aggregated Italy www.epicentro.iss.it/passi/dati/fumo.asp	ISS
Region - No smoking in public spaces - No smoking in work environment - No smoking at home	Second-hand smoking By Region (Italy)- (% population) http://www.epicentro.iss.it/passi/dati/fumoPassivo.asp	ISS

Table 3. Smoking habits by Region (Italy), for each Italian Region and aggregated

Data	Data sources	Provider(s)
Region - % of population	Consumption of fruit & vegetables (5 portions/day) http://www.epicentro.iss.it/passi/dati/frutta.asp	ISS
0 portions: % of population 1-2 portions: % of population 3-4 portions: % of population 5+ portions: % of population	Consumption of fruit & vegetables (5 portions/day) Aggregated Italy http://www.epicentro.iss.it/passi/dati/frutta.asp	ISS

Table 4. Nutritional habits by Region (Italy), for each Italian Region and aggregated

Health data of the population[7]

Data	Data sources	Provider(s)
Region - % of population with diabetes	Incidence of Diabetes by Region Studio PASSI 2011-2014 www.epicentro.iss.it/passi/dati/diabete.asp	ISS

Data	Data sources	Provider(s)
Age range - % of population with diabetes	Incidence of Diabetes Aggregated - Italy Studio PASSI 2011-2014 www.epicentro.iss.it/passi/dati/diabete.asp	ISS
Gender (Male/Female) - % of population with diabetes		
Level of education (No/primary school; secondary school; college; university degree) - % of population with diabetes		
Economic status (high economic difficulty, some economic difficulty, no economic difficulty) - % of population with diabetes		
Citizenship (Italian, Other Country) - % of population with diabetes		

Table 5. Diabetes incidence by Region (Italy), for each Italian Region and aggregated

Data	Data sources	Provider(s)
Region - Hypertension - High cholesterol - At least one CVD risk factor	Cardiovascular risk by Region (% population) Studio PASSI 2011-2014 www.epicentro.iss.it/passi/dati/cardiovascolare.asp	ISS
Age range - % of alcohol consumers - IC95% inferior - IC95% superior	Cardiovascular risk aggregated Italy (% population) Studio PASSI 2011-2014 www.epicentro.iss.it/passi/dati/cardiovascolare.asp	ISS
Gender (Male/Female) - % of alcohol consumers - IC95% inferior - IC95% superior		
Level of education (No/primary school; secondary school; college; university degree) - % of alcohol consumers - IC95% inferior - IC95% superior		

Data	Data sources	Provider(s)
Economic status (high economic difficulty, some economic difficulty, no economic difficulty) - % of alcohol consumers - IC95% inferior - IC95% superior		
Citizenship (Italian, Other Country) - % of alcohol consumers - IC95% inferior - IC95% superior		

Table 6. Cardiovascular risk by Region (Italy), for each Italian Region and aggregated

Data	Data sources	Provider(s)
Region - Perceived health status - Avg. Nb. days bad health (physical+mental) - Avg. Nb. days bad physical health - Avg nb. days with daily activity limitations	Health status (% population) Studio PASSI 2011-2014 http://www.epicentro.iss.it/passi/dati/PercezioneSalute.asp	ISS

Table 7. Perceived health status by Region (Italy), for each Italian Region and aggregated

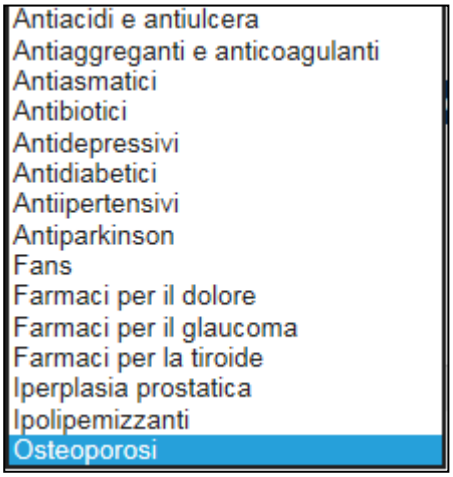
Medications

Data	Data sources	Provider(s)
Medication group[9] - DDD/1000 inhabitants by year	Use of medications by region and medication group - Osservatorio sui farmaci - Italy Period 2000-2011 www.epicentro.iss.it/farmaci/videofar/	Italian NHS, ISS (Cnesps)

Table 8. Use of medications by type of drug - Italy

[9] See Addendum B

Addendum B. medications groups



Hospitalization data

1. Italy: Emilia Romagna Region. Data by district concerning:
   - hospital discharge by gender, age, payer for hospitalization, entity proposing hospital admission: number
   - outpatient visits by gender, age, citizenship: number
   - emergency clinic admissions by gender, age, citizenship: number
   - mortality rate by gender, cause of death,

See: https://applicazioni.regione.emilia-romagna.it/ReportER/. Data can be exported for data analysis.

2. Italy: Lombardy Region. Data by district concerning:
   - hospital admissions by type of intervention: number
   - outpatient visits: number
   - incidence of comorbidity/complicated cases

See: https://www.dati.lombardia.it/browse?category=Sanit%C3%A0&sortBy=relevance&utf8=%E2%9C%93&page=1. Data can be exported for data analysis.

Similar open data repositories are available for many Italian regions.

Environmental data

1. Levels of pollution by region by geographic area (Italy): http://www.viias.it/dataviz/ (interactive maps by air pollutant)
2. Databank of air pollutants from traffic- Italy. Source ISPRA. Consultation by pollutant: http://www.sinanet.isprambiente.it/it/sia-ispra/fetransp/
3. Pollutants emissions years 1990-2014 (Excel file): http://www.sinanet.isprambiente.it/it/sia-ispra/serie-storiche-emissioni/serie-storiche-delle-emissioni-nazionali-di-inquinanti-atmosferici/view
4. Air pollutants (SNAP series) year 1980-2014 - Italy (excel file): http://www.sinanet.isprambiente.it/it/sia-ispra/serie-storiche-emissioni/serie-storiche-delle-emissioni-nazionali-snap/view
5. Environmental performance index – Netherland, Italy (Excel file): http://epi.yale.edu/country/netherlands
6. Organisation for Economic Co-operation and Development: http://www.oecdbetterlifeindex.org/countries/italy/
7. ARPA Emilia Romagna (Excel): http://www.arpa.emr.it/dettaglio_generale.asp?id=2615&idlivello=1521
8. ARPA Lombardia (Excel): http://shp.arpalombardia.it/sites/arpalombardia2013/RSA/Pagine/default.aspx
9. Regione Lombardina (CSV, PDF, XLSX, XML): https://www.dati.lombardia.it/browse?category=Ambiente
10. Netherland Air Pollution facts sheet (Pdf): http://www.eea.europa.eu/themes/air/air-pollution-country-fact-sheets-2014/netherlands-air-pollutant-emissions-country-factsheet
11. Germany Air Pollution facts sheet (Pdf): http://www.eea.europa.eu/themes/air/air-pollution-country-fact-sheets/germany-air-pollutant-emissions-country-factsheet

Public available data and literature

1. Any open source PubMed listed resource relevant for head and neck cancer.

Examples of literature concerning Head and Neck tumor epidemiology:

[1] Gatta G, Botta L, Sánchez MJ, Anderson LA, Pierannunzio D, Licitra L; EUROCARE Working Group. Prognoses and improvement for head and neck cancers diagnosed in Europe in early 2000s: The EUROCARE-5 population-based study. Eur J Cancer. 2015 Sep 6. pii: S0959-8049(15)00749-2. doi: 10.1016/j.ejca.2015.07.043

[2] Van Dijk BA, Gatta G, Capocaccia R, Pierannunzio D, Strojan P, Licitra L; RARECARE Working Group. Rare cancers of the head and neck area in Europe. Eur J Cancer. 2012 Apr;48(6):783-96. doi: 10.1016/j.ejca.2011.08.021. Epub 2011 Nov 1.

[3] Zigon G, Berrino F, Gatta G, Sánchez MJ, van Dijk B, Van Eycken E, Francisci S; EUROCARE Working Group. Prognoses for head and neck cancers in Europe diagnosed in 1995-1999: a population-based study. Ann Oncol. 2011 Jan;22(1):165-74. doi: 10.1093/annonc/mdq306. Epub 2010 Jun 29.

[4] Licitra L, Zigon G, Gatta G, Sánchez MJ, Berrino F; EUROCARE Working Group. Human papillomavirus in HNSCC: a European epidemiologic perspective. Hematol Oncol Clin North Am. 2008 Dec;22(6):1143-53, vii-viii. doi: 10.1016/j.hoc.2008.10.002.

[5] Berrino F, Gatta G. Variation in survival of patients with head and neck cancer in Europe by the site of origin of the tumours. EUROCARE Working Group. Eur J Cancer. 1998 Dec;34(14 Spec No):2154-61.

[6] Survival from salivary glands adenoid cystic carcinoma in European populations. Oral Oncol. 2009 Aug;45(8):669-74.

[7] US Department of Health and Human Services. Cancer and the environment: What you need to know, what you can do.

2. Publications listed at www.epicentro.iss.it

Example: http://www.epicentro.iss.it/igea/diabete/prevalenza.asp