International Journal of Biomedical Data Mining
ISSN: 2090-4924

Vickress et al., Biomedical Data Mining 2014, 3:1
DOI: 10.4172/2090-4924.1000105

Research Article Open Access

Data Inventory for Cancer Patients Receiving Radiotherapy for Outcome Analysis and Modeling

Jason Vickress, Rob Barnett and Slav Yartsev*

London Regional Cancer Program, London Health Sciences Centre, London, ON, N6A 4L6, Canada

Abstract

Objective: To describe a database created for storing and analysis of patient specific data related to pre-treatment condition, treatment planning, and outcomes with a long term future objective to predict the optimal radiation therapy for new patients.

Method: Construction of the centralized database for the collection of sufficient information for outcome analysis and modeling will be comprised of a SQL database and a commercial DICOM-RT PACS (MIM) server. Development of dedicated software for automatic transfer of DICOM-RT files from different sources to MIM PACS through unique import procedures. Planning dose objectives and constraints from Tomoplan (Accuray), Pinnacle (Philips) and Eclipse (Varian) treatment planning systems and daily position correction information from treatment units are transferred to the SQL database.

Results and conclusion: A centralized database for all patient specific data, treatment planning and outcome information allows for determining correlations between treatment parameters and patient outcomes. The proximity between tumor and organs at risk is demonstrated as useful in determining optimal planning parameters in addition to the planning data of previously treated patients. The proposed database can perform automated analysis regarding quality assurance, dose accumulation for multiple treatments on different machines and assist physicians in choosing the optimal treatment modality.

Keywords: Radiation therapy; Personalized medicine; Treatment planning optimization; Treatment modality

Introduction

Data collection for cancer patients is recognized as an important task in the USA, where the National Program of Cancer Registries (NPCR) administered by the Centers for Disease Control and Prevention collects data on the occurrence, type, extent, and location of the cancer, and the type of initial treatment [1]. The International Consortium for Health Outcomes Measurements (ICHOM) aims at providing a global resource of in-use outcome measures and risk adjustment factors by medical condition and creating a global standard for measuring results [2]. These initiatives will enable public health professionals to understand and address the cancer burden more effectively. We have recently proposed to use the pre-treatment, planning, and treatment outcomes data for cancer patients undergoing radiation therapy to provide guidelines for optimal choice of both radiation modality and planning for new patients [3]. It is important to determine the most influential patient features (or their combinations) that has the strongest correlation with the outcomes. We propose an Overlap Volume Histogram as a valuable representation of size and shape for tumor and organs at risk important for planning.

With the advancements in informatics, data collection is done in all aspects of radiotherapy from treatment planning to delivery. These data are spread among individual systems, stored in different formats making information not easily accessible. Each system is designed to work standalone and data accessibility depends heavily on the level of compliance with the IHE-RO standards [4].

Analysis on clinical information from any or all of these different systems requires a significant investment of manpower for data extraction. Several tools have been developed to assist in data analysis with regards to radiotherapy [5-7] but cannot easily integrate all systems together. Here we present a system to collect and record all treatment information for use in analysis and quality assurance. It will involve the storing of DICOM-RT files from multiple Treatment Planning Systems (TPS), treatment parameters and patient specific information (i.e. Outcomes). These data will be made accessible for use by generated tools and scripts for analysis of clinical data and to aid in the planning process of future patients. The ultimate goal of this system is bridging the gap between systems and modalities to get a better picture of the quality of treatment in radiotherapy.

Methods

Commercially available TPSs are all designed to produce the optimized procedure for the delivery device (e.g. linear accelerator or inserted radioactive sources) using patient specific CT study and organ outlines. A dose distribution in the patient is planned that provides the prescribed dose to the tumor while sparing healthy tissues as much as possible. The result of planning is then exported to the treatment unit. However, the TPSs are not designed for data mining, which requires easy export of DICOM-RT files including images, structures and planned dose distribution as well as information about planning process parameters. Overcoming this will require individually constructed processes to have all relevant information exported to a central database.

At the London Regional Cancer Program (LRCP) three different modalities are used for complex external beam radiotherapy:

***Corresponding author:** Slav Yartsev, London Regional Cancer Program, London Health Sciences Centre, London, ON, N6A 4L6, Canada, Tel.: +1-519-685-8600, ext-53171; Fax: +1-519-685-8658; E-mail: slav.yartsev@lhsc.on.ca

Received September 27, 2013; **Accepted** November 27, 2013; **Published** January 14, 2014

Citation: Vickress J, Barnett R, Yartsev S (2014) Data Inventory for Cancer Patients Receiving Radiotherapy for Outcome Analysis and Modeling. Biomedical Data Mining 3: 105. doi: 10.4172/2090-4924.1000105

Citation: Vickress J, Barnett R, Yartsev S (2014) Data Inventory for Cancer Patients Receiving Radiotherapy for Outcome Analysis and Modeling. Biomedical Data Mining 3: 105. doi: 10.4172/2090-4924.1000105

Tomotherapy, static beam Intensity Modulated Radiation Therapy (IMRT) and Volumetric Modulated Arc Therapy (VMAT). Tomotherapy is performed via a TomoTherapy Hi-ART unit (Accuray) and all planning and dose calculations on their proprietary Tomoplan software. Both IMRT and VMAT modalities are performed using a Varian Linac (21× and TrueBeam) with planning and dose calculations done in either Pinnacle (Phillips) or Eclipse (Varian) TPS. All IMRT and VMAT treatments are performed and recorded with the use of an ARIA oncology information system produced by Varian.

For any substantial analysis of relation between patient specific features, planning, and outcomes a number of treatment parameters will need to be available including DICOM-RT files plan specific data, daily treatment changes and patient specific information. Each type of data resides in its own treatment delivery or treatment planning system and requires a unique and automatic export/import method.

DICOM-RT

DICOM-RT files are common in any TPS as they are generated and stored at various stages of the planning process. DICOM-RT files include structures (RS), plan (RP), dose (RD), and images (CT, MRI or PET). For easy access and sharing of all DICOM-RT files they are stored in a central picture archiving and communication system (PACS) server running MIM PACS software (MIM Software Inc. Cleveland, Ohio, USA). A MIM PACS server was chosen because of its PACS connectivity, review features and ability to house DICOM-RT files produced from multiple TPS. The MIM PACS server acts as a clinical storage center for all DICOM-RT information, reducing the need of redundancy in data storage.

For the import of DICOM-RT files into the PACS server each TPS requires a specific procedure. For Tomotherapy it is a part of its semiweekly archiving procedure that includes sending of DICOM-RT files pertaining to each patient at the end of treatment; RS, RP, RD, planning kVCT and daily MVCT image studies. For Pinnacle the MIM PACS server acts as a transfer hub between the TPS and the ARIA treatment database. Normally there is an intermediate PACS server that would facilitate the transfer from the TPS to the treatment unit; MIM PACS is taking over this role. This process requires all patients with completed and approved plans in Pinnacle to be exported to MIM PACS followed by export to the ARIA treatment system. For all patients planned in the Eclipse TPS the DICOM-RT files will be manually exported to the MIM PACS at the end of the treatment cycle as they are already generated within the ARIA information system. This manual process is required for Eclipse because all DICOM-RT files stored in ARIA need to be generated by their tools to be exported in a DICOM-RT format.

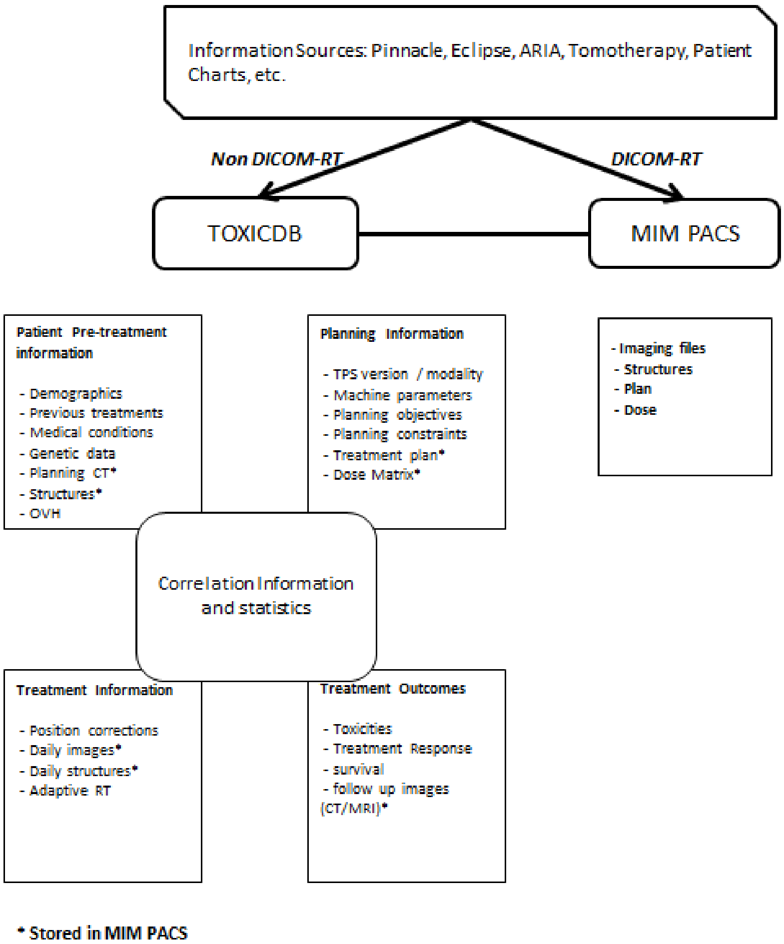

Figure 1: Schematic diagram of the TOXICDB system. Illustrating the integration of MIM PACS with TOXICDB regarding all DICOM-RT files that are physically stored in MIM.

TOXICDB

A separate database titled Treatment-Outcomes-Expanded-Informatic-Correlation Database (TOXICDB) was created to store all non-DICOM-RT information that cannot be managed by MIM PACS in order to facilitate the storage and acquisition of large quantities of diverse information, and maintain the ability to grow in size and diversity. All DICOM-RT files are available from the MIM PACS server, using its PACS connectivity native to all PACS servers. In-house software written to access TOXICDB can retrieve files from MIM PACS with a series of simple C-Find and C-Move commands. Any further information on the connectivity of PACS servers is located in [8]. The DICOM-RT files located with MIM PACS are considered to be a part of the TOXICDB system even though they are physically stored on a different machine. A schematic view of the TOXICDB system is shown in Figure 1, with four separate blocks titled Pre-Treatment, Planning, Treatment and Outcome information.

TOXICDB is a SQL database designed for easy connectivity within a Microsoft environment and with additional software elements connectivity is achieved with Solaris (Unix) for the Pinnacle TPS. None of the specific fields were hard coded into the database to allow further specificity to be detailed in records of the child tables referred to as "Type" tables. The parent tables are capable of storing both numeric and character values. As done in all electronic databases all records are date time stamped, for both creation and modification to maintain the temporal information.

Plan specific data: In radiotherapy the process of creating a unique plan can be a complicated process. Different planning systems use different software and algorithms to construct an effective radiotherapy plan. The planning objectives and constraints are common tools used to construct a plan that meet the physician's requirements. This information is stored within the individual TPS and requires individual extraction methods. For Tomotherapy the objectives and constraints are pulled from the xml files within the archived patient directory named "*_patient.xml". Pinnacle requires the data be exported for each individual plan with execution of specific scripts launched from the Graphical User Interface (GUI). For Eclipse planning, objectives and constraints are located on the ARIA Sybase database and can be accessed by querying the database. All of the objectives and constraints are transferred to TOXICDB for storage and indexing.

A dose objective used for inverse planning is described by three main parameters for each structure of interest; weighting factor (importance), dose-volume constraints, and penalty. The importance is a numeric value that defines the relative impact each structure has on the quality of the plan, typically very vulnerable organs get high weighting, as would the tumor. The dose-volume constraint is the maximum, minimum dose or dose to specific relative volume that the tumor should receive and sensitive tissue tolerate, a penalty value is applied for fine tuning of specific dose-volume objectives. It is a combination of all three parameters that help create unique radiotherapy plans for each patient.

Daily patient position corrections: Both Tomotherapy and Linac are equipped with an integrated image-guidance system (fan-beam MVCT and cone-beam kVCT, respectively) that allows for correction of the patient position using planning CT study and pre-treatment image. These systems calculate recommended shifts to meet the original dose prescription based on the most recent image of the patient. Daily position correction shifts in the lateral, longitudinal, and vertical directions as well as the roll correction are performed with respect to the initial setup made using external marks (tattoos) placed during CT simulation. For each treatment day, a value of the shift in each direction is recorded in the treatment system. For Tomotherapy the correction shifts values are stored within the treatment record and are extracted from the archived patient directories within an xml file "*_patient.xml". For Linacs this information is stored within the ARIA Sybase database in the form of couch positions during imaging and treatment and is retrieved by querying the ARIA database.

Patient Specific Information: Different from the planning and treatment data patient information can come in a number of forms from various sources and includes patient demographics, information from previous treatments, medical history and outcomes. All these

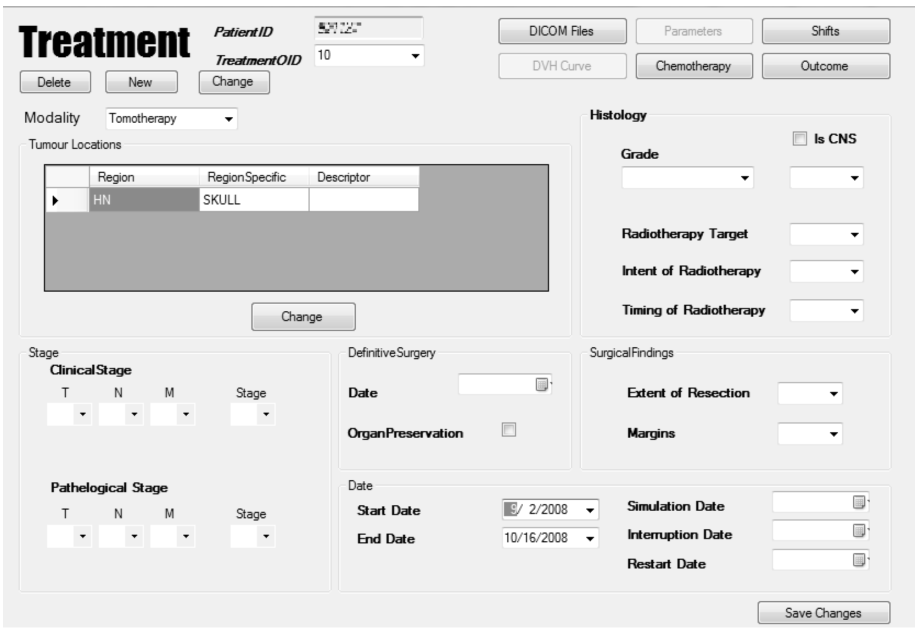

Figure 2: Screen of data import screen for disease and treatment parameters for TOXICDB.

Citation: Vickress J, Barnett R, Yartsev S (2014) Data Inventory for Cancer Patients Receiving Radiotherapy for Outcome Analysis and Modeling. Biomedical Data Mining 3: 105. doi: 10.4172/2090-4924.1000105

data are stored in a centralized TOXICDB database for easy access and retrieval. This information can be collected directly from the patient records or through the perusal of existing clinical trials and imported through an automated process or through a Graphic User Interface (GUI) shown in Figure 2. All fields are linked to specific patients, treatments and date stamped.

All patient specific information is stored and used in compliance with procedures required by the ethics board at LRCP. It is password protected and is contained within the secure clinical servers at LRCP, to maintain patient confidentiality with a possibility of anonymized export.

We included several parameters calculated from DICOM-RT files in order to have an easy description of a patient's anatomy. One of these the Overlap Volume Histogram (OVH) is a parameter used to describe a patient's anatomy with respect to the organs at risk (OAR) and their proximity to the target volume. This parameter represents the minimum distance of every voxel in the OAR to the surface of the tumor volume. The method is a modified version of M. Kazhdan et al. [9] which is described in the equation below

$$OVH(t) = \sum_{p \in O} \{d(p,T) = t\}$$

Where O is the volume of the OAR, T is the surface of the target, and t is a distance from the tumor. The function d(p,T) is a signed distance function which finds the shortest distance between any point in the OAR and a target surface, it is "signed" because it returns a negative value if the point is within the surface of interest. The modification from (M. Kazhdan, et al. 2009) is that the OVH is not normalized by volume, as the volume of each ROI close to the tumor needs to be maintained.

The OVH is calculated using the Structure files obtained remotely from the MIM PACS server and computed using a program coded in C#. The OVH data can be further converted into a cumulative OVH, in the same manner as for a cumulative DVH curve [10]. A cumulative OVH curve represents what volume of a structure is within a certain distance from the target surface and is not currently a supported DICOM-RT structure.

Automation and programming: All automation and programming for TOXICDB are coded using SQL for database commands, and C# for all other functionality. All DICOM functionality is created in C# with the use of two open source libraries titled Fellow-Oak DICOM (fo-dicom) [11] and Evil DICOM [12]. The fo-dicom libraries are used to communicate with the PACS server for access to all DICOM-RT files in all programs and scripts.

Results and Discussion

Automation and scripting

Analysis on DICOM-RT data from a TPS is challenging and requires a case by case approach. Having all the DICOM-RT files located in the MIM PACS server allows for automated analysis. A number of programs have been created to communicate directly with the PACS server to allow the transfer and use of any DICOM-RT files; these were done in C# with the fo-dicom library. This allows a single program to perform analysis on the treatment information of a limitless quantity of patients, with little to no user involvement. With this application the scope of investigation into radiotherapy is greatly increased, with the incorporation of numerous variables.

Quality assurance

Basic quality assurance includes verification of calculations performed by the TPS [13,14]. This requires the use of third party validation using separate systems and measurements [15-17]. Having all the information from each TPS in one location allows for plans reviews and quality assurance (QA) procedures to be performed on each TPS efficiently and with the same protocol.

Cross modality dose accumulation

In radiotherapy it is becoming common practice for patients to have multiple treatments across different modalities. Having all of them stored in one location allows the accumulation of delivered/planned dose from various treatments, using the images, daily positions and plan and dose files [18]. A visual representation of cumulated dose distribution from multiple patient treatments provides necessary information for future courses of treatment as well as evaluation of the effect of accumulated dose. An example of dose accumulation is displayed in Figure 3, with a case of Head and Neck cancer, where the accumulated dose from both treatments was above 70Gy to 80% of the treatment volume. This is especially important for patient re-treatments which typically involve different treatment machines and planning systems.

Cross modality comparison

Determining the optimal treatment modality and TPS for a specific patient can be a difficult task, especially when trying to achieve the best possible treatment [19-21]. With the ability to have information on patients with similar characteristics treated previously with several treatment modalities, any "fudging" each planning system does to favor its own generated plan is removed. The automatic current and retrospective analyses of patients across different modalities determine the relative ability of different treatment modality for a particular

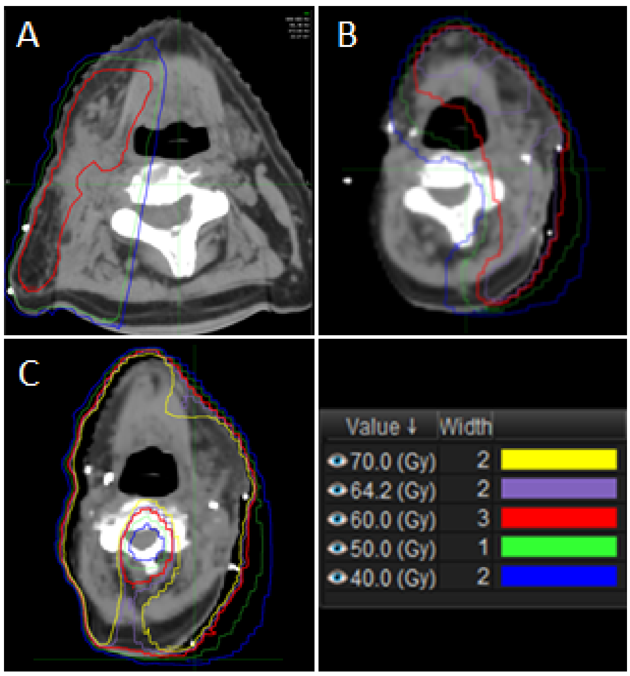

Figure 3: A) CT image with dose distribution of first treatment of IMRT B) CT image with dose distribution of second treatment using Tomotherapy C) Accumulated dose distribution on image B.

patient situation. The information on planning parameters reduces uncertainties and planning time when it comes to the initial steps of the planning process. It also better informs physicians about possible side effects and overall probability of tumor control when using different modalities.

Planning

The planning process depends on the TPS, with different optimization algorithms and level of user input. With the ability to track the planning parameters one can perform an analysis on the quality of the plans, with respect to the parameters chosen. These planning objectives and constraints are usually kept within the TPS because they are not supported by the DICOM-RT structure, but with the suggested system they can be mined from each TPS. With all parameters available one can better understand how to use each TPS to reach optimal treatment plans. Automatic planning, or user guided planning can be integrated with the objectives and constraints stored in TOXICDB [22-25].

Example: overlap volume histogram based comparison and inverse planning

The utility of the TOXICDB system is the combination of multiple data types including treatment and patient specifics across different treatment options in radiotherapy. An OVH is a useful anatomical parameter for evaluation of the correlation between relative positions of OARs and tumor with achievable plan and treatment outcomes.

A computer looks at images in a bitwise fashion making it difficult to view the image as a whole. To overcome this issues an OVH is used to convert a 3D image into a 2D function displaying what volume of a structure is within a distance of the tumor. An example is shown in Figure 4, where comparing these two images directly is very computationally intensive. Using an OVH this comparison can be done much faster, comparing only two 2D datasets. Even though spatial information is lost, this tool can be used for gathering patients of similar anatomical geometries.

The OVH algorithm can also be modified to only consider distances within single slices, or only slices within a certain thickness to account for the width of the field being delivered. Using the OVH data coupled with histology and other patient specific data located in TOXICDB suggestions can be made for treatment based on previous patient's treatments. These suggestions can include planning modalities, constraints and objectives and margins to account for daily changes.

OVH data has also been shown to have uses in the planning process in radiotherapy with respect to determining the correct constraints for the patient [26,27]. In many cases planners determine the starting parameters based on the site being treated (i.e. single brain metastasis). But in many cases the shape and size of the tumor in the same sites can be significantly different and OVH based site similarity becomes more useful.

Guidelines for new patients

The ability to retroactively investigate possible correlations with the use of a database is useful in the evaluation of past treatments. Another application takes the data collected retrospectively to guide decisions for future patients in radiotherapy. This application would go further than current methods using previous data to construct better protocols or procedures for planning and treatment of radiotherapy. The data is directly consulted with the introduction of any new patient using a patient's pre-treatment information including demographics, previous treatments, OVH data etc. to compare them to previously treated patients. The information within TOXICDB provides guidelines made based on successes and failures in the past for patients with similar anatomy and disease. A schematic representation of this process is shown in Figure 5. The final output would be in a form of a report relating certain constraints and objectives and their related side effects. (Figure 5).

Conclusion

Currently in radiotherapy the problem of non-centralized data access is a concern, limiting the number of cross-system analysis that can be performed. The proposed system is designed to connect all of the TPS and treatment delivery systems together for analytical purposes

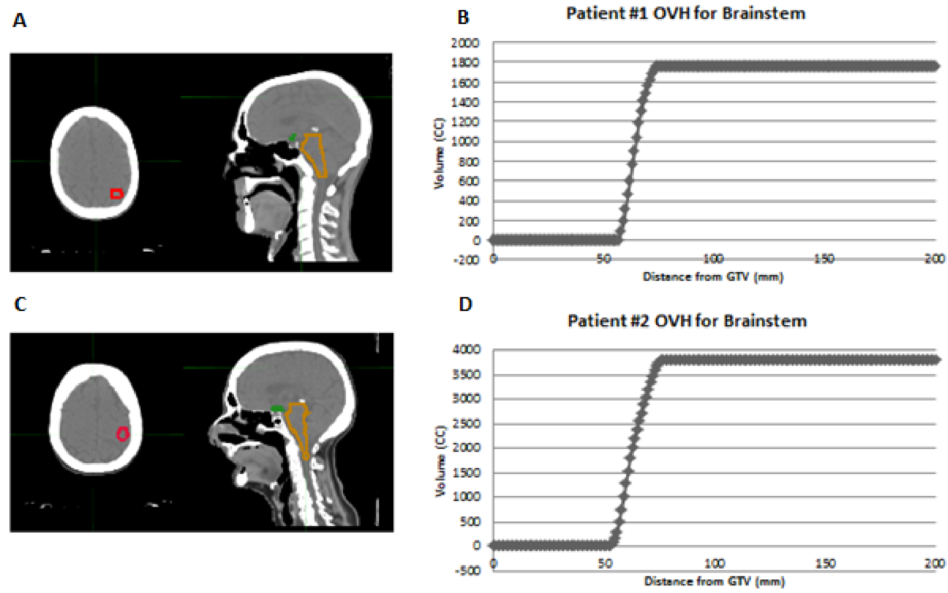

Figure 4: A&C- CT image showing GTV in red, chiasm in green and brainstem in orange. B&D are OVH curves relating the distance between the brainstem and GTV.

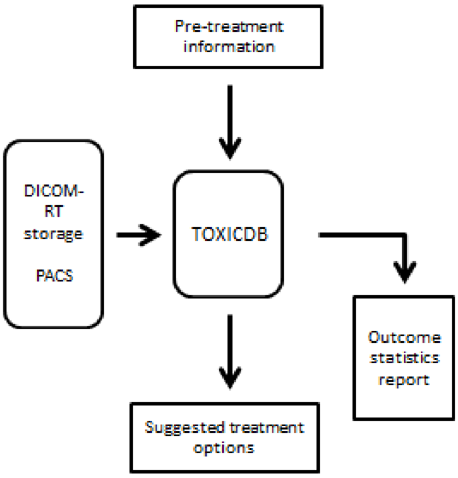

Figure 5: Schematic diagram of TOXICDB based guided planning. For new patients their DICOM-RT and Pre-treatment information is supplied to TOXICDB to generate suggested treatment options and possible outcome statistics.

without impeding clinical procedures. This is done with a series of procedures and automatic programs designed to transfer or re-route information from the parent systems into TOXICDB, through either a SQL database or MIM PACS server. With this centralized data source automatic computerized analysis can be performed efficiently and across all TPSs and modalities. TOXICDB provides full functionality in mixed OS environment. This analysis can be the seed for further developments in radiotherapy treatment and planning.

References

1. National Program of Cancer Registries (NPCR).
2. International Consortium for Health Outcomes Measurements.
3. Yartsev S, Mackie TR (2013) Comment on "When is better best? A multiobjective perspective". Med Phys 40.
4. Abdel-Wahab M, Rengan R, Curran B, Swerdloff S, Miettinen M, et al. (2010) Integrating the healthcare enterprise in radiation oncology plug and play—the future of radiation oncology? Int J Radiat Oncol Biol Phys 76: 333-336.
5. Deasy JO, Blanco AI, Clark VH (2003) CERR: A computational environment for radiotherapy research. Int J Med Phys 30: 979-986.
6. El Naqa I, Suneja G, Lindsay PE, Hope AJ, Alaly JR, et al. (2006) Dose response explorer: an integrated open-source tool for exploring and modelling radiotherapy dose-volume outcome relationships. Phys Med Biol 51: 5719 – 5735.
7. Ebert MA, Haworth A, Kearvell R, Hooton B, Coleman R, et al. (2008) Detailed review and analysis of complex radiotherapy clinical trial planning data: Evaluation and initial experience with the SWAN software system. Radiother Oncol 86: 200-210.
8. Huang HK (2011) Short history of PACS. Part I: USA. Eur J Radiol 78: 163-176.
9. Kazhdan M, Simari P, McNutt T, Wu B, Jacques R, et al. (2009) A shape relationship descriptor for radiation therapy planning, Med Image Comput Comput Assist Interv 12: 100-108.
10. Drzymala RE, Mohan R, Brewster L, Chu J, Goitein M, et al. (1991) Dose-volume histograms. Int J Radiat Oncol Biol Phys 21: 71-78.
11. Fellow Oak Dicom for .Net
12. Evil Dicom by Rex Cardan, Ph.D.
13. Xing L, Currant B, Hill R, Holmes T, Ma L, et al. (1999) Dosimetric verification of a commercial inverse treatment planning system. Phys Med Biol 44: 463-478.
14. Fracescon P, Cora S, Chiovati P (2003) Dose verification of an IMRT treatment planning system with the BEAM EGS4-based Monte Carlo code. Med Phys 30: 144-158.
15. Low DA, Mutic S, Dempsey JF, Gerber RL, Bosch WR, et al. (1998) Quantitative dosimetric verification of an IMRT planning and delivery system. Radiother Oncol 49: 305-316.
16. Low DA, Gerber RL, Mutic S, Purdy JA (1998) Phantoms for IMRT dose distribution measurement and treatment verification. Int J Radiat Oncol Biol Phys 40: 1231-1235.
17. Williams MJ, Metcalfe P (2006) Verification of a rounded leaf-end MLC model used in a radiotherapy treatment planning system. Phys Med Biol 51: N65-N78.
18. Janssens G, Xivry JO, Fekkes S, Dekker A, Macq B, et al. (2009) Evaluation of nonrigid registration models for interfraction dose accumulation in radiotherapy. Med Phys 36: 4268-4277.
19. Xia P, Fu KK, Wong GW, Akazawa C, Verhey LJ (2000) Comparison of treatment plans involving intensity-modulated radiotherapy for nasopharyngeal carcinoma. Int J Radiat Oncol Biol Phys 48: 329-337.
20. Mohan R, Barest G, Brewster LJ, Chui CS, Kutcher GJ, et al. (1988) A comprehensive three-dimensional radiation treatment planning system. Int J Radiat Oncol Biol Phys 15: 481-495.
21. Ezzell GA (1996) Genetic and geometric optimization of three-dimensional radiation therapy treatment planning. Med Phys 23: 293-306.
22. Purdie TG, Dinniwell RE, Letourneau D, Hill C, Sharpe MB (2011) Automated Planning of Tangential Breast Intensity-Modulated Radiotherapy Using Heuristic Optimization. Int J Radiat Oncol Biol Phys 81: 575-583.
23. Reinstein LE, Wang XH, Burman CM, Chen Z, Mohan R, et al. (1998) A Feasibility Study of Automated Inverse Treatment Planning for Cancer of the Prostate. Int J Radiat Oncol Biol Phys 40: 207-214.
24. Ketting CH, Austin-Seymour M, Kalet I, Jacky J, Kromhout-schiro S, et al. (1997) Automated planning target volume generation: An evaluation pitting a computer-based tool against human experts. Int J Radiat Oncol Biol Phys 37: 697-704.
25. Woudstra E, Heijmen BJ (2003) Automated beam angle and weight selection in radiotherapy treatment planning applied to pancreas tumors. Int J Radiat Oncol Biol Phys 56: 878-888.
26. Wu B, Ricchetti F, Sanguineti G, Kazhdan M, Simari P, et al. (2011) Data-driven approach to generating achievable dose-volume histogram objectives in intensity-modulated radiotherapy. Int J Radiat Oncol Biol Phys 79: 1241-1247.
27. Wu B, Pang D, Simari P, Taylor R, Sanguineti G, et al. (2013) Using overlap volume histogram and IMRT plan data to guide and automate VMAT planning: A head-and-neck case study, Med Phys 40.