

Special Section on C&G VCBM

Uncertainty guidance in proton therapy planning visualization

Maath Musleh^{a,*}, Ludvig Paul Muren^{b,c}, Laura Toussaint^{b,c}, Anne Vestergaard^b,
Eduard Gröller^{a,d}, Renata G. Raidou^a

^a TU Wien, Favoritenstrasse 9-11/E193-2, 1040, Vienna, Austria

^b Danish Centre for Particle Therapy, Aarhus University Hospital, Nørrebrogade 44, 8000 Aarhus C, Denmark

^c Department of Clinical Medicine, Aarhus University, Palle Juul-Jensens Boulevard 99 8200 Aarhus N, Denmark

^d VRVis Research Center, Donau-City-Straße 11, 1220 Wien, Austria

ARTICLE INFO

Article history:

Received 28 October 2022

Received in revised form 14 February 2023

Accepted 15 February 2023

Available online 20 February 2023

Keywords:

Visual analytics

Applied computing

Decision support systems

ABSTRACT

We investigate uncertainty guidance mechanisms to support proton therapy (PT) planning visualization. Uncertainties in the PT workflow pose significant challenges for navigating treatment plan data and selecting the most optimal plan among alternatives. Although guidance techniques have not yet been applied to PT planning scenarios, they have successfully supported sense- and decision-making processes in other contexts. We hypothesize that augmenting PT uncertainty visualization with guidance may influence the intended users' perceived confidence and provide new insights. To this end, we follow an iterative co-design process with domain experts to develop a visualization dashboard enhanced with distinct level-of-detail uncertainty guidance mechanisms. Our approach classifies uncertainty guidance into two dimensions: *degree of intrusiveness* and *detail-orientation*. Our dashboard supports the comparison of multiple treatment plans (i.e., nominal plans with their translational variations) while accounting for multiple uncertainty factors. We subsequently evaluate the designed and developed strategies by assessing perceived confidence and effectiveness during a sense- and decision-making process. Our findings indicate that uncertainty guidance in PT planning visualization does not necessarily impact the perceived confidence of the users in the process. Nonetheless, it provides new insights and raises uncertainty awareness during treatment plan selection. This observation was particularly evident for users with longer experience in PT planning.

© 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Proton therapy (PT) is a standard radiation modality in cancer treatment. It requires careful planning to ensure that a tumor will be sufficiently irradiated while adjacent tissues are avoided as much as possible. The treatment plan is calculated in a dedicated treatment planning software (TPS), which computes how the therapy system will deliver the radiation dose to the patient. As this is a lengthy process, it is usually limited to generating a couple of alternatives with additional positional (i.e., translational) variations. Deciding on an *optimal plan* is a complex undertaking with several uncertainty factors, which relate to the physics behind the calculations and the biological effects of the dose on tissues.

Researchers and practitioners designing and selecting robust PT plans for patients depend on the available TPS to calculate the

plan(s) and make appropriate therapy decisions. The TPS includes mainly slice-dose overlay views (Fig. 1(a)) and an additional plot, called *dose volume histogram* (DVH), that depicts radiation dose administered to volume percentages of specific structures (Fig. 1(b)). Beyond juxtaposition, the TPS views do not allow the user to simultaneously compare and assess multiple plans. They also do not support understanding the involved PT uncertainty factors and how these might affect the treatment outcome. The current workflow leaves a gap for a more efficient sense- and decision-making process, exploiting the synergy between the human expertise of the researchers or practitioners in PT, and the computational power of the TPS.

Guidance techniques [1,2] could leverage the synergy between domain experts and their systems, but they have yet to be explored within the context of PT or other clinical applications. In clinical applications (and thus also in PT), data exploration through visual interfaces is often complex. There is still a significant lack of *trust* in visual analysis frameworks and *confidence* in the outcomes [3]. It is reflected by the low adoption of visualization solutions in clinical workflows, which indicates that the suitability of visualization frameworks for clinical decision-making scenarios is limited [3]. In PT, uncertainty adds to the

* Corresponding author.

E-mail addresses: maath.musleh@tuwien.ac.at (M. Musleh),

ludvmure@rn.dk (L.P. Muren), lautou@rn.dk (L. Toussaint), annveste@rn.dk (A. Vestergaard), groeller@cgt.tuwien.ac.at (E. Gröller), rraidou@cgt.tuwien.ac.at (R.G. Raidou).

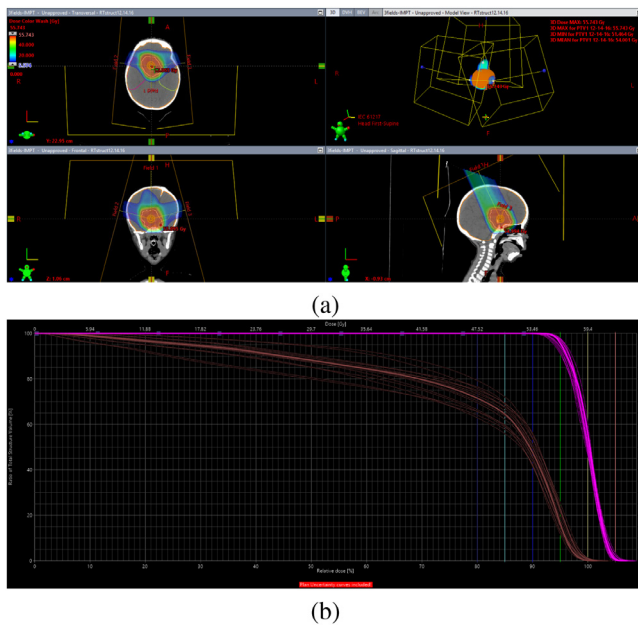


Fig. 1. Conventional views in available treatment planning systems (TPS): (a) slice-dose overlay views and (b) dose volume histogram (DVH), where brown indicates the brain stem (a structure at risk) and magenta indicates the tumor target. The line band corresponds to the plan's robustness.

sense- and decision-making complexity [4]. Developing *guidance techniques for uncertainty* within a PT planning visualization system is anticipated to provide researchers and practitioners with a comprehensive view of the planning robustness. Hence, we aim to provide an overview of the entire decision space, which includes all involved uncertainties and their impact on the treatment plan. Our approach is expected to improve the effectiveness and reproducibility of the sense- and decision-making processes in PT planning.

The *contribution* of this work is the design, development, and assessment of a dashboard as a guided visual interface that enables: (i) the effective comparison of PT plans and (ii) the analysis of the impact of their respective uncertainties. As part of the interface, we investigate and develop *guidance mechanisms* that facilitate navigating through PT plans and their uncertainties in a *multi-level-of-detail manner*, where both the degree of intrusiveness and detail-orientation of the guidance can be tuned and adapted to the needs of the user. Our proposed two-dimensional guidance mechanism (i.e., intrusiveness and detail-orientation) has not been addressed before, while guidance has yet to be investigated extensively in the context of a clinical application. To *evaluate guidance* in PT uncertainty visualization, we propose a framework that focuses on the impact of the guidance on the sense- and decision-making process.

2. Background

2.1. Clinical background

PT requires careful planning to account for several uncertainty factors—from data acquisition to treatment planning and radiation dose delivery. In the planning stage, researchers and practitioners use dedicated TPS, such as Eclipse [5]. They calculate possible treatment plans by deciding on characteristics of the treatment, such as the number of beams, their directions, or specific dose constraints to be fulfilled. Given the computational complexity of the process, only a few (i.e., 2–3) *nominal plans*

are computed. Within the few nominal plans, practitioners also consider several uncertainty factors in deciding which of the plans is the best solution for a patient. Therefore, each treatment plan alternative encompasses a large possible combination of uncertainties [6]. Currently, some uncertainties are accounted for during treatment planning through *robust optimization*.

Despite the significant effort in the visualization community to formalize the definition of uncertainty [7,8], our field has yet to adopt a standard definition. For radiotherapy, Raidou defines *uncertainty* as “any variation in the dose planning outcome, which is produced by an ad-hoc choice or a stochastic process at any stage of the radiotherapy pipeline” [4]. This definition outlines the challenging and multi-faceted nature of uncertainty in radiotherapy (and, subsequently, in PT). In common practice, not all sources of uncertainty are addressed at once. Several sources of uncertainties are prioritized due to their high impact on deciding the final treatment plan. For example, previous work focused only on uncertainty due to anatomical variability [9,10]. Here, we address other types of uncertainty—related to the underlying physics and biological effects of the dose on tissue and their subsequent side effects.

Set-up uncertainties are often investigated in the domain of radiotherapy. They represent the possible deviations in patient positioning on the treatment couch compared to the expected position from which the plan is initially calculated [11]. These deviations, in practice, reflect positional variations along the three main anatomical axes of the patient. For example, for each nominal plan, six additional variations along the $\pm x$, $\pm y$, and $\pm z$ axes of the patient are considered. Additional uncertainties are related to the *relative biological effectiveness (RBE)* [12], which is an emerging uncertainty topic in PT that is not yet tackled in TPS. Clinicians use a fixed RBE factor of 1.1 to account for the larger effectiveness of proton- compared to the photon-based RT [13, 14]. However, several mathematical models exist that provide a more accurate factor [15]. These calculations also encompass another level of uncertainty with the α/β sensitivity variable per structure, which is essential for the RBE value [16,17].

We schematically depict in Fig. 2 the whole process, which results in several factors and alternatives for each nominal plan. These need to be considered by clinical physicists to make an informed and complete decision about an optimal plan to follow. Currently, these alternatives are explored and analyzed manually using slice-dose overlay views, DVH plots, or both (Fig. 1). This solution does not offer a comprehensive view of the different PT plan alternatives' sense- and decision-making space. It also does not support clinical physicists in their workflow of investigating PT plan uncertainties to decide on a robust treatment strategy for a given patient.

2.2. Important concepts and definitions

Here, we clarify essential concepts and terms in the paper:

Clinical Goals: Clinical guidelines provide a list of dose exposure limits for each structure based on scientific evidence. This list is used as a reference when physicians inspect dose exposure, and domain experts refer to them as *clinical goals*. Clinical goals help them identify a therapy decision that avoids unwanted damage to structures.

Guidance: Guidance is ambiguously discussed in visualization and visual analytics (VA), and many different definitions (and interpretations thereof) exist. Guidance comes in many forms: some are prominent (e.g., text popping up to provide suggestions), and others are subtle (e.g., visual cues) [2]. In this work, we adopt the definition of Ceneda et al. [18] that defines *guidance* as “a computer-assisted process that aims to actively resolve a knowledge gap encountered by users during an interactive visual

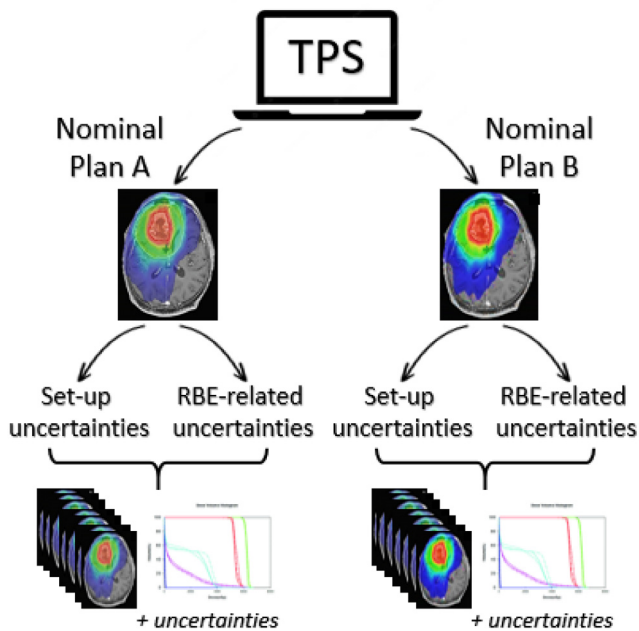


Fig. 2. Schematic depiction of the uncertainties involved in deciding on an optimal PT plan. For each nominal plan, set-up and RBE-related uncertainties have to be accounted for.

analytics session”. In our case, the computer-assisted process of analyzing PT plans and their respective uncertainties is facilitated by a two-dimensional mechanism, i.e., intrusiveness vs. detail-orientation, which can be tuned and adapted to user’s needs. This mechanism enables the adaptive use of (visual) cues derived from the data and presented in our interactive visual system to bridge the users’ knowledge gap and inform their analysis. We employ guidance as a means to unveil the impact of uncertainties on the decision-making space and to raise awareness about this impact.

Guidance Intrusiveness: By intrusiveness, we define the degree to which guidance interjects a sense- and decision-making process. It can be *high* if the guidance amends or supports existing visual plots by, e.g., line-styling or *low* when a separate, auxiliary view presents the guidance.

Guidance Detail-Oriented: By detail-orientation, we define the detail level, or data resolution, at which guidance is applied. In our case, it can be applied per voxel, per structure of interest, or per slice—depending on the level where the analytical process of the user is occurring.

Knowledge Gap: By knowledge gap, we define the quantifiable difference between the required knowledge to complete an analysis task and the knowledge so far obtained by the user. A gap may arise from different types of knowledge, such as domain knowledge or VA tool knowledge [19]. In the context of our work, the knowledge gap arises from lacking an overview on the entire radiotherapy planning decision space, which is hindered by multi-sourced PT plan uncertainties. Referring to the categorization of Ceneda et al. [20], the *type* of our knowledge gap relates to an *unknown target* and pertains to the *data domain*, as the workflow revolves around understanding data uncertainties and their impact on the simulated dose data. Based on discussions with domain experts, this knowledge gap cannot only be identified in the interaction with the tool and during the analytical process [20]—and even before the start of the analytical process. By integrating guidance cues in our tool, we preemptively support the users in their sense- and decision-making process and provide them with a complete view of the information required to make an informed decision on the optimal plan.

3. Related work

Guidance in Visualization—Ceneda et al. [20] describe guidance methods in visualization. They characterize three aspects to provide a systematic framework for developing guidance in visualization systems: knowledge gap, input and output, and guidance degree. Their framework gives a concise perspective concerning guidance in visualization. According to their understanding, a guidance technique is only relevant in visualization if a specific knowledge gap is identified in the interaction with the tool. This gap is bridged with methods that lead users through the process. Guidance is “not merely an additional algorithm that computes results”. They conclude that there is a *lack of guidance techniques that accompany the user in the entire process*. An extension is proposed to understand the guidance approach’s input. They also claim the need for guidance techniques that provide multiple output forms.

To further clarify the concept of guidance, Ceneda et al. [1] suggest a decision tree to determine whether a synergy between humans and computers would provide a more satisfactory outcome. This decision mechanism on using guidance helps to think about the technique itself. However, the decision tree assumes a singular target per visualization system. In reality, the main objective of a visualization system is a consolidation of several targets. An open question is whether a guidance technique should be considered per target or system.

In follow-up work, Ceneda et al. [2] review guidance approaches currently popular in visualization. They investigate a dimension that helps to position guidance work. Our work is located on the axis of system guidance that supports data exploration and the verification and generation of new knowledge about the data. In the context of data uncertainty-related problems, an orienting guidance approach is preferable, i.e., one that *supports maintaining mental maps*.

Sperrle et al. [21] examines a co-adaptive learning model to expand on the mixed-initiative process model introduced by Ceneda et al. [2]. The work examines how user and system knowledge converge through a co-guidance mechanism to bridge the *knowledge gap*. The co-adaptation aims “to reach a high degree of machine automation”. Nonetheless, Ceneda et al. [18] emphasize that guidance intends to provide a suggestion rather than “close the knowledge gap automatically”. As the system reaches the automation stage described in the co-adaptation model, the automated actions of the system cannot be classified as guidance anymore. The co-adaptive model introduces an interesting perspective to develop a dynamic guidance technique that learns from the analytical provenance.

Stoiber et al. [22] distinguish guidance techniques from onboarding. An intrusive form of step-by-step guidance described in Stoiber’s work would be resisted in PT planning. In their definition of guidance, an ambiguity persists on what qualifies as guidance in visualization. Their definition differs from Ceneda et al.’s [18] emphasis that guidance does not compute results; instead, it is “a catalyst for human–computer interaction”. Nonetheless, Stoiber et al. [22,23] propose several aspects that could be employed in developing guidance, such as *the questions of where and how to provide guidance*.

Several works also look into applications where the use of a sensitivity value is disputed. In exploring the finance domain, Torsney-Weir et al. [24] enabled users to manipulate a sensitivity variable while exploring the data. This feature led to an increase in trust in the visualization. They highlight in their work the distraction possibilities in case of intrusive features. Furthermore, expert users might want to ignore the feature. A complementary finding by Bögl et al. [25] suggests that domain experts tend to accept a novel visualization if it relies heavily on conventional

visual encodings. This conclusion is especially applicable in the medical field and PT planning as well [6], where several models are used to calculate dose uncertainties.

In an earlier paper, Torsney-Weir et al. [26] outline decision-making strategies in uncertainty visualization. The six strategies provide a reasonable basis for considering visual encoding and design decisions. The study of Torsney-Weir et al. identifies the lexicographic decision-making strategy as being very popular in the visualization community. It largely corresponds to the decision-making process involved in PT planning visualization. The clinician decides to prioritize certain aspects to make the final treatment decision [27]. For example, in exploring radiation doses to a brain tumor, the clinician prioritizes whether structures around the brain stem and other regions of interest (ROI) are spared or harmed.

Guidance for Uncertainty Visualization—Uncertainty visualizations for biomedical applications follow the general trends in the parent field of uncertainty visualization, despite their additional specific characteristics [8]. Approaches to visualize uncertain data from environmental, ecology, and urban studies [28–32] could also be transferred to biomedical applications. We refer to the survey of Schlachter et al. about existing work in uncertainty visualization for the radiotherapy domain [27]. It also demonstrated that an explicit and thorough investigation of uncertainty guidance principles in PT planning visualization is still missing. Nevertheless, there is work on guidance for uncertainty visualization in other contexts, which we consider relevant for our work.

Belyakov et al. [33] propose a guidance approach to address uncertainty in spatial data analysis in cartography. They study their visual encoding approaches in terms of *context* in their work. Given the data and uncertainty characteristics, such an approach could be translated from cartography to the uncertainty domain in PT visualization. The basis of PT planning visualization is mapping dose data to medical images. This resembles the use of scalar and spatial data in cartography.

Florice et al. [34] also rely on context to decide on the visual encodings employed in their system THALIS. They position their work on the axis of knowledge discovery. Their visualizations aim to minimize uncertainties in the symptom data of cancer patients. THALIS does not qualify as a guidance technique based on the definition in our paper. It utilizes machine learning to present results rather than guide users through decision-making. However, the visual encoding used to reduce the data complexity and uncertainty is transferable to the field of guidance for uncertainty scenarios [35].

Finally, Kamal et al. [36] surveyed visualization approaches to represent uncertainty in data. They conclude that a comparative visualization approach is required to represent uncertainty accurately. Moreover, they suggest giving users control over when and which uncertainties are encoded, hinting at level-of-detail approaches.

The work we surveyed highlights a gap in the literature regarding visualization techniques that incorporate multiple levels of uncertainty in their visual interfaces. Despite efforts to characterize guidance techniques, systematic approaches still need to be included regarding the development and application of uncertainty guidance mechanisms. This is also the case for the medical visualization domain. Moreover, the literature does not present concrete evaluation methodologies that measure user confidence—a critical aim in tackling uncertainties. Our work addresses this gap in the state of the art.

4. Task analysis and research questions

After several rounds of discussions among the co-authors of this paper (visualization researchers and medical physicists), we jointly agreed that the current PT workflow is missing two main aspects. First, medical physicists need visualization mechanisms that support them in *comparing and assessing plan alternatives* resulting from the process described in Section 2 and depicted in Fig. 2 (T1). Second, the workflow should enable the guided exploration of the *impact of different uncertainty types (set-up and RBE) on the final decision* about the optimal plan (T2). Making sense of the entire decision space is a complex process requiring a complete awareness of many uncertainty sources. To unveil how the decision space is affected by uncertainties, guidance is required [19].

Clinical applications often face the problem of *adoptability*. They might be using very complex and unfamiliar visual representations (or interfaces, in general), trying to heavily alter the existing workflow of the domain experts, or both. To counteract this, we provide only simple views that are already familiar to the intended users, and we do not enforce guidance through the analytical workflow.

As our co-designing medical physicists remarked, the workflow should be *tunable to their analytical needs*. The users should be able to decide the degree of aid the system should provide when comparing and assessing plan alternatives and their uncertainties. Tunability, i.e., *control*, is an important requirement for clinical users that empowers them to reach a confident decision on the perceived optimal therapy plan. We use guidance to unveil the impact of uncertainties on the decision-making space and raise awareness about this impact. Nevertheless, the final decision has to be made by the domain expert in a way that is consistent and compatible with the current workflow.

We investigate the following *research questions*:

- RQ1** How can guidance assist *PT plan comparison* (T1) and *reduce the uncertainty complexity* (T2)?
- RQ2** Which guidance methods can improve the *confidence* of users in PT plan sense- and decision-making processes?
- RQ3** How can we *validate guidance techniques* designed for uncertainty PT visualization?

To address **RQ1** and **RQ2**, we design a guidance dashboard specific to the scenario of PT uncertainty visualization. It supports the exploration and analysis of multiple PT plans (T1) along with their uncertainties and their impact on the decision-making process (T2). Only simple representations are employed to support *adoptability*. Specifically for **RQ2**, we investigate different levels of intrusiveness and detail in developing and applying our guidance mechanisms. This approach is necessary to provide the desired level of *control*. Finally, the value of our adopted methods is measured through a domain-expert user evaluation. It builds upon available approaches to account for the guidance impact on the sense- and decision-making process **RQ3**.

5. Design and implementation

Our designed and implemented dashboard supports comparing PT plan alternatives. The guided navigation of their uncertainties uses a multi-level-of-detail approach with different degrees of intrusiveness. The dashboard is developed in a flexible and extensible PyCharm environment using *Python* and *Dash*. It is the result of an iterative co-design process between visualization researchers and clinical physicists, as further discussed in Section 6.1.

5.1. Patient data and PT uncertainties

We received anonymized planning Computed Tomography (pCT) data with two sets of PT plans from pediatric brain tumor patients. The 3D plans were calculated for illustrative purposes in the context of this work. These included two nominal plans and six different set-up uncertainty scenarios for each nominal plan. In total, we have 14 plan alternatives per patient resulting from the set-up uncertainties. The alternatives are all 3D volumes, where the scalars encode radiation dose in Gray (Gy).

We obtained the respective linear energy transfer (LET) distributions for each of the plans. These are also 3D scalar volumes, where a scalar encodes the amount of energy an ionizing particle transfers to the material through a unit distance traversal in $\text{keV}/\mu\text{m}$. The data represent the energy deposition density along the depth of the beam, which is an essential variable in estimating damages caused by the radiation dose to the tissues [37]. The LET is also an input parameter for the different RBE models proposed in the literature.

Moreover, the datasets included a delineation of a total of 93 structures located around/close to the tumor, such as the brain stem or the temporal lobe, and target delineations. We also received a list of maximum-dose clinical goals (maximum radiation dose) as input to inform guidance in calculating the DVHs for the different structures.

In addition to the set-up uncertainties, we deal with RBE-related uncertainties, which need to be calculated in real time and integrated into our dashboard. We calculate the RBE values using the LET distribution and eight known models in the field [15]. Thus, we obtain eight RBE calculations for each plan, which may differ from each other. Additionally, we use a default structure sensitivity value α/β of 2.0 Gy, but the interface enables the user to control these values in real time while comparing PT plans. This functionality enables the user to unravel the RBE uncertainty further.

5.2. Uncertainty guidance design

We propose a *two-dimensional formulation of uncertainty guidance*, which targets complex scenarios where a multi-level-of-detail – yet controlled – exploration of uncertainties is required. Table 1 shows an overview and categorization of the guidance mechanisms developed as part of our dashboard and further discussed in Section 5.3.

The first dimension represents *intrusiveness*. As discussed in Section 2.2, *low-degree intrusiveness* does not “intrude” the user’s analytical visual space; instead, it employs on-demand, auxiliary solutions. *High-degree intrusiveness* becomes intrinsic to the user’s analytical visual space by altering it. Our definition indicates that highly intrusive guidance is displayed within the user’s direct perceptual focus or is integrated into conventional views, such as *new encodings within* the slice–dose overlay or the DVH view. Oppositely, less intrusive guidance comes within separate, supplementary views that can be added on-demand. This salient encoding is intended to draw the user’s attention proactively. Intrusiveness relates – but is not entirely equivalent – to the controllability characteristic of guidance, as discussed by Ceneda et al. [19]. The system allows the user to control the guidance degree and, subsequently, the intrusiveness degree to comply with their analytical process. This feature also resembles the *refine* stage, as described by Sperrle et al. [21]. As a difference, our approach does not include automated control; instead, it leaves tuning to the user.

The second dimension represents *detail-orientation*, which relates to the level of detail (i.e., voxel, structure, or slice level) at which the guidance works. At each of these levels, we investigate

and compare the impact of uncertainties on the patient plan. The structure level is the most common one, as planning often considers entire organs [6]. The voxel level provides more granular detail on sub-parts of the organ where significant differences might occur. The slice level is an aggregation essential for efficiently comparing doses and identifying issues at very impactful slice ranges, e.g., close to tumors or structures.

In designing our guidance mechanisms, we consider additional points. First, we integrate uncertainty guidance with anatomical and DVH plots, maintaining *conventions* in the PT domain to promote adoptability. Second, we employ the guidance plots as *abstract visual cues*. We abstract the plots if low-level details are not meaningful for the analytical process of the user. For example, if we are interested in a high-level comparison of two or more dose distributions, we largely omit axis labels and ticks as unnecessary detail. This way, we accommodate many plans and uncertainty factors in an abstract comparative view to provide indications of plan differences that guide the user through the decision-making process. Finally, our guidance techniques encompass *visual organization techniques*. Thus, we do not introduce any new plots. We only rearrange and regroup existing ones. The following subsection discusses how we implemented the guidance-enriched PT planning dashboard.

5.3. Dashboard implementation

We followed a detailed-context-overview approach while building our dashboard [38]. The final dashboard evolved through an iterative co-design process (further described in Section 6). It resulted in the inclusion of different views that enables the users to explore and compare details and summaries of plan-dose overlays with their uncertainties. The dashboard also provides several control features to enable dynamic guidance based on user interaction.

As *input*, the dashboard requires DICOM (Digital Imaging and Communications in Medicine) data that comprise a patient’s pCT scan slices, information about the structures of interest (delineations and dose-limitation clinical goals), and the respective dose and LET distributions. As *output*, the dashboard displays a comparative view of different PT plans and their uncertainties (RBE and set-up). Concerning the *available knowledge*, the domain experts are knowledgeable and experienced in reading and comprehending dose/LET plan visualizations overlaid on pCT scans together with structure contours and DVH plots (Fig. 1). The domain experts are also knowledgeable about the nature of uncertainties. However, they could benefit from visualizations that support them in understanding and comparing uncertainties from different plan alternatives (T1,2). In this case, conventional plan visualizations and DVH plots are insufficient, as discussed in Section 2. This is, thus, the *knowledge gap* we attempt to bridge.

Slice–Dose Overlay View—Fig. 3 includes the typical slice–dose overlay view on the pCT slices along the three main anatomical axes for two plans. This slice-based view is conventionally used in PT to represent the dose distribution on the patient’s anatomy at each voxel position [27]. In the background, the pCT scan is presented using a grayscale, and the dose- (or LET-) plan is overlaid using a rainbow color map. Additionally, structure contours are overlaid with distinct colors. The rainbow color map and the contour colors are retained to follow standard practices and conventions in PT visualization [27]. This view supports T1. To optimize space on the interface for the comparison, we position the three anatomical planes vertically, as opposed to the configuration in Fig. 1(a). A 3D view was not deemed necessary by our collaborators.

To compare two plans, the domain experts currently must exchange the view by moving back and forth between plans.

Table 1

Guidance overview and categorization using two dimensions (intrusiveness and detail-orientation) and employed in our dashboard to compare the available PT plans and navigate through their respective uncertainties.

		Detail-orientation		
		Per voxel	Per structure	Per slice
Intrusiveness	Low	Slice–dose overlay (Fig. 3) & Explicit comparison view (Fig. 4)	RBE uncertainty violin plot (Fig. 6)	Uncertainty indicator (Fig. 8)
	High	On-demand uncertainty distribution plot (Fig. 5)	Targeted stylization of DVH lines for structures of interest (Fig. 7)	Re-adjustment of the presented dose based on selected RBE models

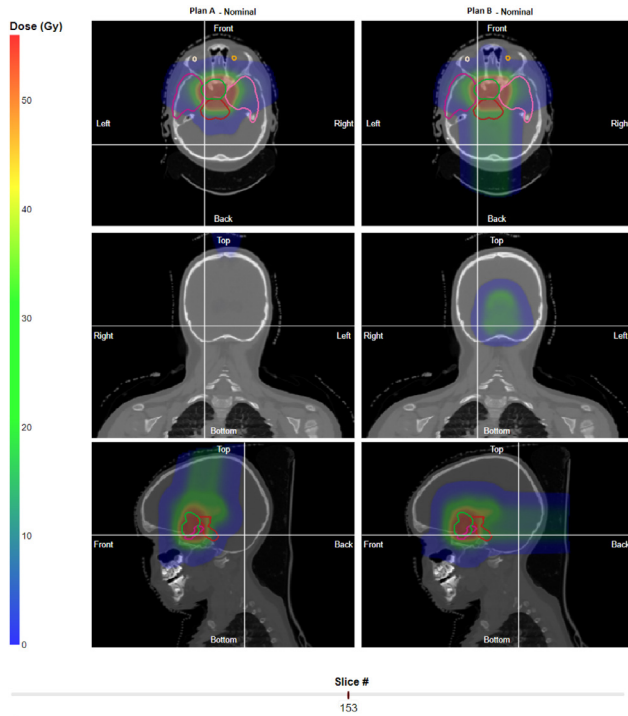


Fig. 3. pCT slices overlaid with dose distributions and structure contours from two plans (in the two columns) on the three main anatomical planes (in the three rows) of the patient for one slice (Slice 153). The contour colors follow the conventions in the clinical environment.

Plumlee and Ware [39] suggest multiple windows to compare complex data that the users cannot easily hold in their visual working memory. In our dashboard, we adopt a juxtaposition approach where the two plans are depicted side-by-side. This supports plan comparison through a simultaneous and linked exploration of the slices. It is a *low-intrusiveness* and *per-voxel* guidance mechanism, which allows a one-by-one comparison of two dose- (or LET-) plans for one patient at a voxel level. Superposition is not an adequate choice, as the underlying planning CT scan has to be retained for anatomical context.

Explicit Comparison View—For a more explicit comparison of two plans, we use the view shown in Fig. 4. Explicitly encoding of computed difference provides better precision of the image difference [40], which supports the comprehensibility demand of a guidance encoding. The view encodes the difference between the primary (top) plane views in Fig. 3. It includes the typical anatomical slice with a semitransparent dose overlay that results from the absolute difference between the dose or LET values of the first and the second plan. A calculated value may be positive or negative, based on which plan receives a higher dose at the specific voxel position. The difference is encoded in a red-to-blue diverging color scale. Red represents a higher dose value

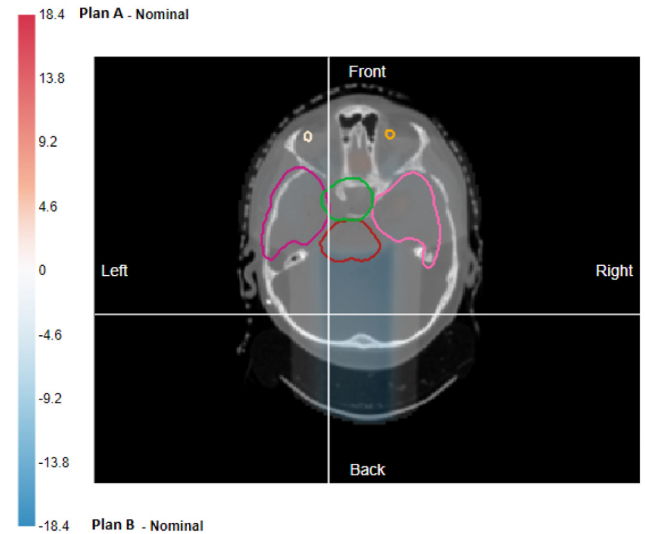


Fig. 4. The explicit comparison view encodes the differences between the two plans of Fig. 3 using a divergent red-to-blue colormap to indicate different doses between the two alternatives at each voxel. The contour colors follow the conventions in the clinical environment.

for the first plan, and blue represents a higher dose value for the second at a given voxel position. The view can additionally include structure delineations, as used conventionally in the domain of PT planning. It implements a *low-intrusiveness* and *per-voxel* guidance, which supplements the slice–dose overlay view. The explicit comparison view supports **T1**.

Uncertainty Distribution Plot—The user can hover over a voxel position in an ROI that receives a radiation dose (e.g., a voxel within the brain stem). In this case, a plot of the dose distributions of the alternative plans at that specific voxel with accompanying uncertainties will appear. Several comparison techniques are possible. Due to the high-intrusiveness nature of this encoding, we use superposition to provide a direct visualization of the trade-off between the two plans [40].

The plot compares at a high level and highlights differences in the distribution of all possible dose values at a voxel. It is based on all different RBE and set-up uncertainty factors for each plan and structure (Fig. 5), and supporting **T2**. We depict each plan with a distinct color (red or blue, following the explicit comparison convention). The rugplot [41] under the horizontal axis further enhances this representation by visualizing the marginal distribution of the data as marks along the axis. Rugplots are often coupled with distribution plots to enhance the view on the raw data used in plotting the distribution. This guidance mechanism provides users with a many-to-many robustness indication in comparing two plans at a voxel level with a *high degree of intrusiveness*.

RBE Uncertainty Violin Plot—The view in Fig. 6 provides a guidance of *low degree of intrusiveness* to the user on the robustness of plans. Robustness is based on the uncertainty distribution

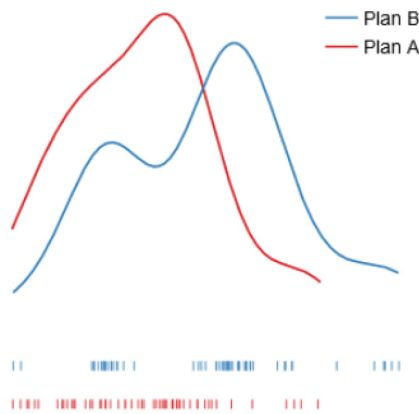


Fig. 5. The distribution plot pops up when hovering over a voxel position within a delineated structure receiving a radiation dose. It reveals the high-level uncertainty distributions of the two plans indicated with the two different colors. The rugplot, i.e., the marks on the horizontal axis, further enhance the abstract comparative representation.

introduced by the calculated RBE factor of each structure of interest (T2). We represent the RBE uncertainty distribution for each structure of interest with a split violin plot, as seen in Fig. 6. It displays at a high level the distribution of the possible values based on calculating RBE at each scenario of the plan with each of the chosen calculation model(s) and other parametrizations for the structures of interest.

We considered alternatives like common box plots or sparse representations, such as those proposed by Wentzel et al. [42]. Although box plots might be more familiar to domain experts, violin plots are advantageous for showing the entire data distribution [43]. They also facilitate the identification of differences across distributions. This characteristic enables the users to detect marginal trade-offs between the compared plans, which is crucial in our scenario. The user can zoom into areas of interest. Additional statistical annotations are provided on-demand to support the comparison of the two distributions—showing also exact numerical values. Eight models are used to calculate the RBE factors from each set-up scenario at each voxel position (in total, 56 different alternatives) for each structure and both plans. In Fig. 6, a subset thereof is shown.

Targeted Stylization of Dose Volume Histogram (DVH) Lines—The DVH plot is a typical depiction in the RT domain that displays the distribution of the dose over volume percentages of structures (see Fig. 1(b)). The distribution of dose values to volume percentages is calculated from the DICOM data files, which contain the delineations of the structures of interest. In the DVH plot, users can inspect the dose or LET volume distribution. DVH plots accommodate a many-to-many per-structure comparison (T1,2). Different plans and uncertainty factors can be computed and plotted with color coding and different line styles. In Fig. 7, we illustrate two plans in the same DVH plot with several structures of interest. The user selects and deselects structures for each plan through the active legend. Line stylization supports an intrusive guiding strategy. A distribution line is drawn with a larger width if the corresponding structure received a maximum dose that exceeds the clinical goals. In Fig. 7, this occurs for the temporal lobes in plans A and B. Adjusting the line width provides a more expressive visual guidance cue, given the lightness of the conventional colors used to represent the structures. A less intrusive mechanism may involve uncertainty bands [44], where the per-structure set-up uncertainty is displayed as a band around the line.

Uncertainty Indicator—A heatmap plot summarizes the plan robustness per slice, i.e., which plan has less uncertainty for each

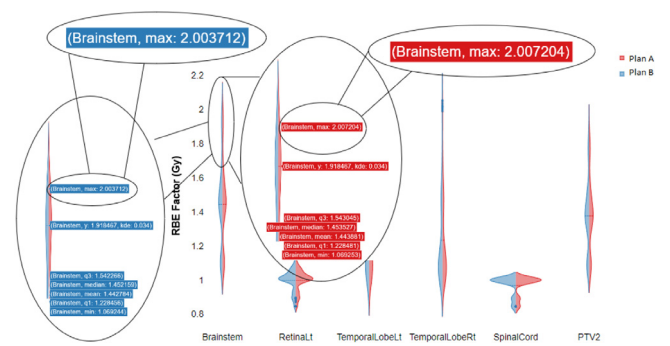


Fig. 6. Comparative uncertainty distribution of possible RBE factors for the structures of interest. The two plans are shown with two distinct colors. The RBE uncertainty calculations include many factors, such as different RBE models and structure sensitivities (α/β ratio). When hovering, the user is provided with additional annotations to support the comparison. The calculations have been done for six structures of interest.

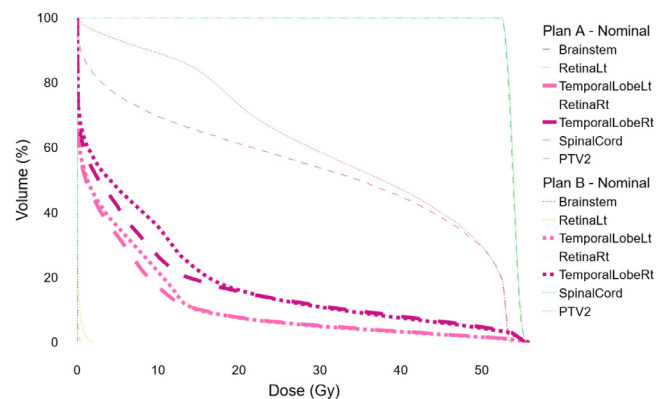


Fig. 7. The Dose Volume Histogram (DVH) for the structures involved in two plans (plan A indicated with dashed lines and plan B with dotted lines). Line stylization guides the user in identifying structures at risk (encoded by the increased line width, e.g., for the temporal lobes in plans A and B). The colors have been selected based on TPS conventions.

slice. This view supports T2. We use a grayscale heatmap to avoid user confusion, given that the dashboard already employs other color encodings dictated by field conventions.

We calculate the average dose exposure and standard deviation for each slice, which resulted from the different RBE calculation models. The standard deviation σ indicates the variability in the uncertainties throughout all possible alternatives. σ is encoded in a grayscale where dark gray indicates a high σ , i.e., a slice with high uncertainty. An example of the uncertainty indicator view is shown in Fig. 8, juxtaposed are two plans and the different slices of the volumes. In this example, slices beyond number 230 have high σ s, i.e., uncertainty, in plan A and low σ s in plan B. Brushing and linking allow the user to select a particular slice by directly clicking on the uncertainty indicator to investigate the slice further. The view integrates a low-intrusiveness guidance to support uncertainty plan comparison at a slice level.

Re-adjustment of the Presented Dose based on Selected RBE Models—The user can select which models to include in calculating the average RBE factor per voxel. It is done through several controls, further described below. Changing the model will trigger a real-time recalculation and adjustment of the views affected by changes in the RBE computation. It can be experienced as a high-intrusiveness solution, mainly applicable to slice-based exploration and analysis. The re-adjustment mechanism supports T1 and T2.

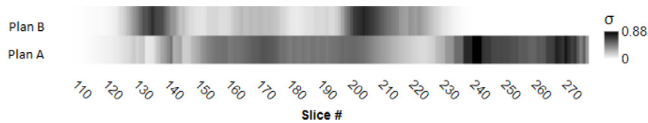


Fig. 8. The uncertainty indicator view provides a per-slice indication of the uncertainty for two plans and all slices. A darker gray value indicates a higher σ , i.e., higher uncertainty, in the given slice.

Controls—The dashboard is supported through several controls that accompany the guidance strategies. Some controls are shared between views, and others are restricted to an individual view. In total, the dashboard includes five controls, one for: manipulating the slice-based view, the inclusion/exclusion of structures of interest, setting up the clinical goals and sensitivity value for each structure (Fig. 9(a)), exchanging the RBE models and set-up scenarios (Fig. 9(b)), selecting the level of guidance intrusiveness that the user prefers (Fig. 9(c)).

Making a selection through one of the controls has a linked impact on one or more views of the dashboard. For example, the controls of Fig. 9(a) provide an input mechanism for the user to include or exclude structures and set their sensitivity values and clinical goals. Modifications directly impact dose calculation, which is then reflected on the slice-dose overlay view, the explicit comparison view, and, subsequently, on all the other views of the dashboard. Similarly, the controls of Fig. 9(b) concern the inclusion of different RBE uncertainty sources and set-up uncertainties, affecting all possible dashboard views.

Finally, the slider in Fig. 9(c) controls the degree of guidance intrusiveness in the dashboard views. The user can decide how intrusive the uncertainty guidance should be in her sense- and decision-making process. The control starts from disabled guidance, where the conventional slice-based views in Fig. 3 and the DVH plot in Fig. 7 are presented. In the minimal and intermediate degrees of intrusiveness, low-intrusive guidance is added to the interface. In the former, the explicit comparison view (Fig. 4) is included, and in the latter, the RBE uncertainty violin plot (Fig. 6) and uncertainty indicator (Fig. 8) are added. The full guidance mode includes the intrusive guidance that styles lines in the DVH plot (Fig. 7) and the uncertainty distribution plot on slice pixel hovering (Fig. 3). It also presents the RBE calculation controls in Fig. 9(a) and (b).

5.4. Usage scenario

In this section, we present a usage scenario to showcase the full functionality of our dashboard. We expect the users to start their navigation with the lowest degree of guidance, as this coincides with current practices. Initially, the slider under the panels is used to browse the linked anatomical slices in Fig. 3. The users scroll through the data until they reach a slice that displays the region(s) of interest (ROI). For example, in Fig. 3, these could be the temporal lobes (indicated in pink and magenta), brain stem (indicated in red), and tumor volume (indicated in green). They can, subsequently, add more guidance and consult with the explicit comparison view in Fig. 4. Here, they can identify a significant difference in dose radiation in specific ROIs (T1), i.e., the posterior part of the brain receives a higher radiation dose with plan B (indicated by the blue color). In Fig. 9(a), the users can deselect the contours of the least important regions. Hovering over a structure allows them to explore the uncertainty in the voxel values, e.g., in Fig. 5, which shows a shift between the two plan distributions (T2). To receive a more comprehensive picture, they can change the displayed data or select a higher dose

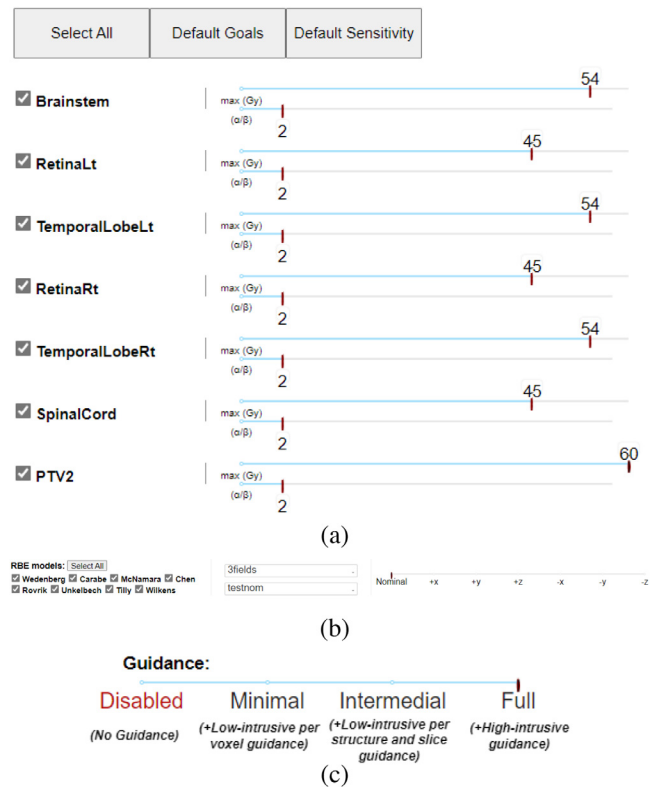


Fig. 9. Three of the five controls accompanying the guidance mechanisms of our dashboard: (a) provides a check-list of the structures of interest and sliders for setting their clinical goals and sensitivity values, (b) provides a check-list of the included RBE models, a drop-down menu to choose plans for comparison, and a slider for selecting set-up uncertainty scenarios, and (c) provides an interface to select the level of guidance intrusiveness.

value range to display. Users inspect the plot in Fig. 8 to identify particular slices where a significant uncertainty variance occurs. In this example, they can select slice 138, where plan A exhibits a significant uncertainty variance indicated with a darker color. Slice 138 is also depicted in Fig. 4, where a difference in dose is noticeable.

To complement their findings, the users can opt for the highest degree of guidance and request to additionally explore the DVH plot in Fig. 7. In this example, the increased line thickness indicates structures that may be at risk (T1,2). For instance, the right temporal lobe indicated with the dashed, thick, magenta line might be a structure at risk both in plans A and B. In contrast to this, the retina (light pink line) is not a structure at risk in either of the two plans. The users can then customize the clinical goal for an interesting structure at risk and deselect the rest. They navigate the differences between the dose distributions in the structure in both plans using the plot shown in Fig. 5. Here, the two distributions differ by a shift, indicating that plan B administers a higher radiation dose overall. To get a clearer picture of the plan robustness at a structure level, users can further explore the violin plots in Fig. 6. The plots may indicate significant differences between the plans for each of the structures. In this example, the differences are slight, as the two sides of the violin plot are almost identical. Finally, users can also customize the sensitivity values and RBE calculation models in Fig. 9(a) and (b), respectively, and include the sensitivity uncertainties in the analysis. Now, users can make decisions relying on a high or low degree of guidance intrusiveness in our tool.

Table 2
Domain experts participating in the evaluation sessions and their experience (in years) in RT research and practice.

	Occupation	Research years	Practice years
P1	Medical Physicist	25	5
P2	PostDoc Researcher	5	0
P3	Medical Physicist	19	24
P4	PhD Student	1	0
P5	Medical Doctor	6	3
P6	Medical Physicist	18	30

Table 3
Participation of domain experts in the different stages of the evaluation (formative sessions vs. user evaluation).

	Evaluation stage			User evaluation
	Formative sessions			
	1	2	3	
P1	✓	✓	✓	
P2	✓	✓	✓	✓
P3	✓			✓
P4				✓
P5				✓
P6				✓

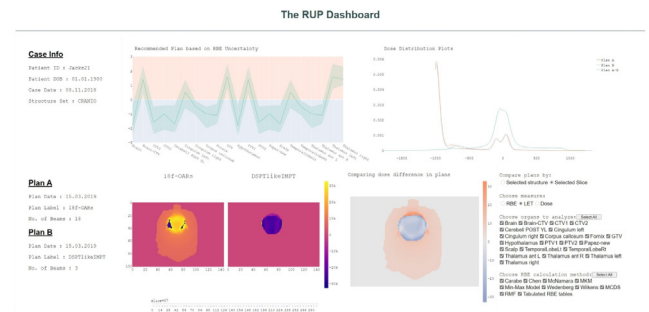
6. Evaluation and results

We followed an iterative co-design process, where the views and user interactions discussed in Section 5 were developed in tight collaboration with domain experts. This approach has facilitated understanding problems in the PT planning domain [45]. The iterative design included *three formative sessions* to inform our design decisions (Section 6.1). After the finalization of the design, we conducted a *user evaluation* (Section 6.2). The aim was to evaluate whether the dashboard, with its guidance mechanisms at different degrees of intrusiveness, influences the perceived confidence and supports the users in their sense- and decision-making process. We included in total six domain experts (Table 2) throughout the formative sessions and the user evaluation (Table 3).

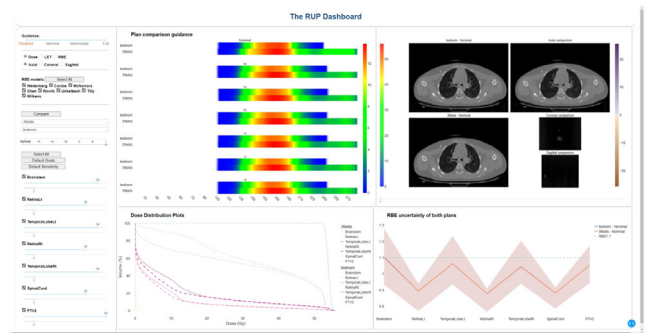
6.1. Formative evaluation

We conducted three formative evaluation sessions to inform our design decisions. As seen in Table 3, P1 and P2 were involved in the entire co-design process. They initiated this project by providing data, initial information regarding current practices and the clinical workflow, as well as ideas for improving the current sense- and decision-making process.

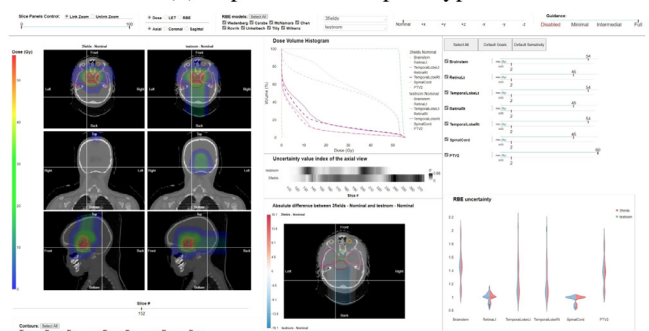
The *first formative session* was conducted together with P1, P2, and P3 in a joint session. We discussed in this session a low-fidelity prototype sketch (Fig. 10(a)) to elaborate further on how the users imagine the design of the dashboard. We discussed *which tasks users would like to be able to perform* and *how their current workflow could be improved*. We also dealt with *visual cues* and how we could design them better to ensure *compliance* with field conventions. P1 and P2 highlighted the unresolved visualization of uncertainties in the process as their current knowledge gap. Presently, the field of PT is discussing the possibility to select one or two RBE calculation models instead of the 1.1 RBE factor, which is the current clinical default value. After this session, we improved the comparison encoding in the RBE uncertainty violin plot based on discussions with the expert collaborators. We also condensed the slice views to provide a permanent display of the three plane views—axial, sagittal, and coronal. Moreover, we included domain-familiar visualizations, i.e., the DVH plot and slice views. Finally, we added controls to enable user customization



(a) Low-fidelity prototype



(b) Improved second prototype



(c) Final dashboard

Fig. 10. The evolution of our dashboard through the three formative evaluation sessions. (a) Low-fidelity prototype used in the first session. (b) Interactive improved prototype used in the second session. (c) Final dashboard employed in the user evaluation.

and a heatmap plot to indicate plan robustness. It resulted in the second design (Fig. 10(b)).

The *second formative session* was conducted individually with P1 and P2 to discuss the second design iteration. This session was targeted toward *concretizing the design of the individual views*. We started by discussing the views and functionalities of the dashboard. Then, P1 and P2 obtained remote control and freely explored the dashboard. We received valuable feedback regarding functionality and visual cues to inform our third design iteration. After this session, we changed the parallel coordinates plot to a violin plot to represent more accurately the RBE uncertainty. We rearranged the views based on feedback about the anticipated analysis process. Furthermore, we made the slice views larger and permanently displayed the three slice views of both plans in a juxtaposed manner to better support comparisons. Finally, we redesigned the heatmap plot into an aggregated version. It resulted in the third design (Fig. 10(c)).

In the *third formative session*, we discussed the improved iteration of our dashboard. The purpose of this session was to *receive final feedback from both co-designers* before putting the

Table 4
The 7-point Likert scale questionnaire of the ICE+SUS (inspired by previous works [46,47]), completed by the study participants.

Insight	Strongly Disagree						Strongly Agree
1. The visualization helps me detect and understand the differences between the two plans	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
2. The comparison between plans is difficult	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
3. The visualization provides different level-of-detail information when exploring the data	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
4. The information provided by the panels in the dashboard is redundant	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Confidence							
5. The visualization reduces the complexity of the data	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
6. The visualization uses meaningful and accurate visual representations	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
7. The visualization is a misleading representation of the data	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
8. The knowledge presented on the screen is complicated	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Essence							
9. The system does not provide more value than existing tools	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
10. The visualization helps me make a decision about a plan recommendation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
11. The visualization provides an understanding of the uncertainties in the data	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
12. The visual representation of uncertainty in the data did not impact my decision	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

dashboard out for the main user evaluation with a larger group, which is discussed in Section 6.2. As illustrated in Fig. 10, the dashboard evolved through several stages of an iterative process, where together with domain experts we identified design choices appropriate to fulfill their tasks. Some views were included in the first two versions of the dashboard (e.g., parallel coordinates to show uncertainties, or different comparative views), but did not make it to the final dashboard, as they were not considered effective or insightful enough.

6.2. User evaluation

User Evaluation Design—In our study, we evaluate the impact of guidance techniques on the sense- and decision-making process. We also assess the potential of the uncertainty visualization approaches in reducing the perceived complexity in the calculated PT plans. Specifically, we aimed to answer the following questions:

EQ1 How does the user employ the dashboard with regard to the given uncertainty guidance mechanisms at different levels of intrusiveness (low vs. high)?

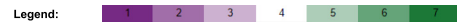
EQ2 How does the user employ the dashboard with regard to the given encodings at different detail-orientations (voxel, structure, and slice)?

To answer the questions, the user evaluation design was informed by the taxonomy of scenarios presented by Lam et al. [48]. We evaluate first the *visual data analysis and reasoning (VDAR) scenario*. The study participants explore real-patient data in a controlled environment while we observe their behavior in employing the dashboard. Moreover, we wanted to assess how the guidance mechanisms affected *user performance (UP)* in exploring the data and making a decision. For this, we used an evaluation framework, which we called ICE+SUS, as it was inspired by both the ICE-T questionnaire by Wall et al. [46] and the System Usability Study (SUS) popularized by John Brooke [47]. In this questionnaire (Table 4), six statements were worded negatively, and the remaining six were positive statements to minimize acquiescence and extreme response biases. The participants were requested to fill in this questionnaire after using the dashboard with different guidance levels. Furthermore, we evaluated *user experience (UE)* by assessing whether the encodings effectively reduced perceived uncertainties and increased the users' confidence. For this, the responses to the ICE+SUS survey and answers to additional open-ended questions were solicited. The open-ended questions regard what could be improved, reworked, or seen as useful among the uncertainty guidance mechanisms of the dashboard.

User Evaluation Course—The final dashboard, as resulting from the formative sessions (Fig. 10(c)), underwent an online

Table 5
ICE+SUS feedback results from the three sub-sessions of the user evaluation: With full guidance (high intrusiveness), with intermediate guidance, and without guidance (low intrusiveness). Point 1 indicates the strongest disagreement, and 7 denotes the strongest agreement.

Full-guidance				
	Insight	Confidence	Essence	Average
Participant 2	6.00	5.00	4.75	5.25
Participant 3	6.25	4.75	5.75	5.58
Participant 4	5.25	4.50	4.75	4.83
Participant 5	4.50	4.25	4.50	4.42
Participant 6	6.00	3.25	4.25	4.50
95% C.I.	5.60 ± 0.89	4.35 ± 0.84	4.80 ± 0.71	4.92 ± 0.62
Intermediate Guidance				
	Insight	Confidence	Essence	Average
Participant 2	5.75	5.25	4.75	5.25
Participant 3	4.25	4.75	5.75	4.92
Participant 4	4.25	4.50	6.00	4.92
Participant 5	4.50	2.75	3.25	3.50
Participant 6	5.75	3.75	4.00	4.50
95% C.I.	4.90 ± 0.97	4.20 ± 1.21	4.75 ± 1.44	4.62 ± 0.84
No Guidance				
	Insight	Confidence	Essence	Average
Participant 2	4.25	4.50	3.25	4.00
Participant 3	2.25	6.00	5.00	4.42
Participant 4	5.50	5.00	4.50	5.00
Participant 5	4.75	2.75	2.50	3.33
Participant 6	4.25	4.75	3.50	4.17
95% C.I.	4.20 ± 1.49	4.60 ± 1.47	3.75 ± 1.24	4.18 ± 0.75



evaluation with five participants (P2–P6). P1 did not participate due to extensive familiarization with the dashboard. We conducted individual online and recorded sessions using screen sharing and remote control. We spent the first few minutes explaining the views of the interface and basic functionalities of the dashboard until these were clear to each participant. Then, the participant was given control to freely explore the dashboard without our interference.

The user evaluation was divided into three sub-sessions, where the intrusiveness level of the guidance mechanisms was increasing. To control for bias, we randomized the order of the sub-sessions. At the end of each sub-session, the users made a recommendation on the plans and filled out the questionnaire (ICE+SUS and the open-ended questions) to address the UE and UP scenarios. The results of this part are discussed in the remainder of this section. The VDAR part resulted in use cases; one example is described in Section 6.3.

User Evaluation Outcomes—The results of the study with regard to the intrusiveness dimension (EQ1) are presented in Table 5. They show a constant improvement with increasing intrusiveness in the guidance mechanisms, in terms of *gained insights* and *essence*, i.e., added value to the current workflow. This improvement was constant across all participants, except for P4, who provided the most positive feedback for low intrusiveness approaches. A reason might be in P4 being the least experienced among our study participants (see Table 2). On the other side, participants with the highest RT research and practice experiences,

i.e., P3 and P6, commented that they gained the least insight with the no-guidance dashboard. Their insights increased when presented with full guidance. However, we noticed a decrease in *reported confidence* as more intrusive guidance was introduced to the interface. The two highly-experienced participants (P3 and P6, Table 2) also expressed less confidence in the uncertainty encoding. The results may imply that, in the current state, the dashboard does not increase the perceived confidence of the users. However, it provides them with significantly more insight through the fully guided interface and adds value to their current workflows. The confidence issue could also be due to a lack of familiarity with the newly introduced visual representations. It is worthwhile to investigate this aspect in a longer-term field study.

With regard to detail-orientation (EQ2), we observed mixed behaviors. The participants had very different preferences regarding views, encodings, and detail-orientation. The slice-based views and the DVH plot seem to be the most helpful to the process based on all participants' feedback. It is “*a matter of habits*”, according to P2. P5 made use of the uncertainty distribution pop-up in the slice-based view, and the styling of the DVH lines guided her. Moreover, the uncertainty indicator seemed confusing to the participants as they struggled to “*grasp what was actually represented*” (P2). P4 thought that the RBE violin plot was difficult to read, although it could be useful to make decisions at a structure level. Most participants agreed that seeing the RBE uncertainty range is useful to their decision-making process. However, P3 believes that using one model at a time for the calculation is a better approach. On the other hand, P6 was skeptical about “*accept[ing] RBE calculations clinically*” overall. P6 was the only participant to believe that the new views in the dashboard (beyond the conventional ones) would not be helpful to the decision-making process. The other participants agreed that, given complete RBE calculation results, the dashboard could provide new insights into the sense- and decision-making processes.

Finally, we observed the users' decision-making processes during the study. Seemingly, the different views of the system were well integrated as users navigated the data at different detail levels. As anticipated, for an initial analysis, the participants were over-reliant on conventional views, shown in Figs. 3 and 7. Nonetheless, they employed the intrusive guidance provided in these views, shown in Fig. 5 and explained in Fig. 7. They regularly inspected the violin plot in Fig. 6, which provided them with a comparative view of RBE uncertainty at a structure level, and the slice plot in Fig. 4, which highlighted areas of difference at a slice and pixel level. On the other hand, the uncertainty indicator view (Fig. 8), which provides quick guidance of interesting uncertainty deviations in slices, was not used efficiently. The users expressed confusion over how the plot derives its values, which highlighted the potential need for guidance explainability.

6.3. Use case example

In this section, we present a use case example. It has been conducted by one of our evaluation participants (P5, medical doctor) as part of the VDAR scenario. Further cases were conducted with the same patient data, and different participants analyzed their cases following different analytical processes and using different guidance levels and detail-orientations. The case below is just an example; no clinical inferences should be derived from it.

Participant P5 chooses the two plans to compare with the dashboard. The brain stem is the most important structure to be spared. She hovers over its contour voxels to explore the dose value it receives. The uncertainty distribution comparison pops up. The plot provides per-voxel insight into the difference in robustness of the two plans. Plan B (Fig. 11(a) in blue) shows larger uncertainty. Subsequently, she looks at the comparison plot

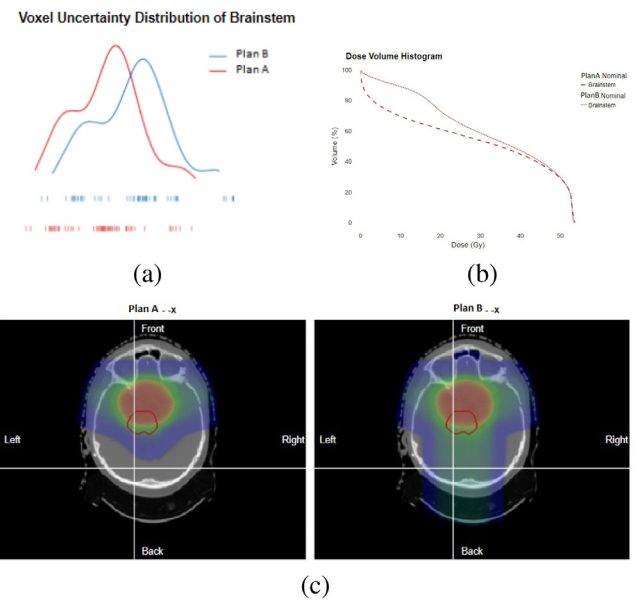


Fig. 11. Comparison of two plans as conducted in a use case by P5, using the functionality of our dashboard. The contour colors follow the conventions in the clinical environment.

and sees an area of interest (per-structure analysis). She zooms into the brain stem region for a closer look and sets a customized clinical goal for the maximally accepted dose (54 Gy). She notices, in Fig. 11(b), the DVH line thickening, and she understands that the structure has received a higher dose than it should. She further inspects the dose distribution for both plans (per-slice analysis). She zooms in to the highest dose to see which plan is “*further out beyond 54 Gy*” and explores different set-up scenarios. She has indications that plan A could be more robust and may incur fewer risks to the brain stem, as seen in the slice analysis of Fig. 11(c). Plan A might be recommended in this case over plan B. The case example indicates that participant P5 has employed all levels of detail (per voxel, structure, and slice) but not all views (i.e., the RBE violin plot and the uncertainty indicator were not used). Higher intrusiveness guidance was mostly preferred in her sense- and decision-making process (and later positively assessed, as seen in Table 5).

7. Discussion

In this section, we outline the lessons learned from the feedback received in the user evaluation and during the uncertainty guidance dashboard design. We designed a visualization dashboard supported with uncertainty guidance mechanisms to improve the users' confidence in their sense- and decision-making process in PT planning, as expressed in Section 4. The dashboard supports two tasks: the comparison and assessment of multiple PT plans (T1) and the analysis of their respective uncertainties, along with their impact on the treatment outcomes (T2). It is the first attempt to integrate different detail levels of guidance and degrees of intrusiveness to a clinical sense- and decision-making process. We propose an evaluation framework to successfully measure confidence in guidance techniques by adapting two evaluation schemes to the needs of our application (ICE-T and SUS).

This work highlights the literature gap in developing usable visualization tools that provide guidance to users of PT planning (and other clinical) systems. The design process indicates that guidance can be a suitable strategy for supporting PT plan comparison. It *may generate more insights and add value to the current*

workflows (RQ1). It is particularly the case for more experienced users, e.g., P3 and P6, who seem to prefer increased levels of guidance, i.e., highly intrusive guidance. In Table 5 P3 and P6 report increasing average points of 1.16 and 0.33, respectively, when using the full-guidance dashboard in comparison to using it without guidance. Regarding detail-orientation guidance, the prior analytical preferences of the users seem to matter.

The user evaluation also shows that *obtaining more insights does not necessarily mean increased user confidence (RQ2)*. Although P3 reports an increase of 4 points in insight obtained, it corresponded to a reported 1.25 points decrease in confidence. The reduced confidence could be attributed to familiarization with the conventional decision-making process. At the same time, all detail-orientation levels seem to be needed, even though different users might have different preferences on the views to use. Observing the sessions, we noticed that the participants relied more on conventional tools, such as those shown in Figs. 3 and 7, in exploring the therapy plans.

Participants in the user evaluation expressed interest in the comparative violin plots shown in Fig. 6. The view enabled them to compare an essential dimension of uncertainties in plans, i.e., RBE values at a structure level, which was not visible to them before our work. This feature seems to be appreciated highly in PT decision-making. P3 reported that the tool is “*useful to get an overview of how the RBE uncertainties for critical organs compare*”. However, guidance explainability could boost users’ trust in the tool. To increase the value of the technique, an explainable guidance for conveying the underlying calculations to the users would be a helpful future addition and increase confidence in the tool.

The variance in the participants’ experiences, as outlined in Table 2, contributed to the high variance in the evaluation results. Highly experienced domain experts might be more skeptical about the underlying uncertainty calculations or about adopting new approaches in their workflows. We recommend involving participants with varying experience levels throughout all design phases. From the elaborate feedback at the end of the sessions, we receive encouraging indications that our approach has value for PT planning. Additional studies with complete PT plan data are required to develop a more mature approach that could be adopted in clinical practice. *Onboarding techniques could be helpful in this case, as well as long-term studies (RQ3)*.

The main limitation in developing and evaluating the dashboard was the restriction to use a synthetic dataset that cannot accurately account for more uncertainty components. Additional LET data could provide a more realistic evaluation setup, as the currently limited LET data affects the RBE uncertainty calculations and the final PT plan decision. Domain conventions and expectations contributed to a skeptical attitude toward the introduced dashboard. A further developed prototype or longer training could contribute to addressing this issue. It could highlight the visualization techniques’ value as complementary to conventional processes rather than being antagonistic. In addition, some of the representations in our dashboard have been designed to strictly follow conventions in the domain. However, the design could be improved and evaluated in the future according to best practices in the visualization domain (e.g., colormaps for the dose overlay or the contours and encodings in the DVH plots). Also, analyzing multiple datasets, i.e., from different patients or tumor localizations, would be valuable.

Finally, our observation is that guidance approaches fit very well with provenance approaches [49], which can provide additional knowledge on the path which leads to a decision. The proposed two-dimensional guidance concept can be generalized to other domains. We reckon that the approach applies to three areas in particular: weather forecasting, financial prediction, and material science. The data in these three domains embed multi-level uncertainties, which need to be explored at different detail levels, similar to PT.

8. Conclusions and future work

We presented a two-dimensional uncertainty guidance to support comparison and uncertainty analysis in PT planning visualization. We exemplify uncertainty guidance on a PT planning dashboard for plan comparison. Moreover, an evaluation measured confidence and insights from the use of uncertainty guidance. The evaluation yielded encouraging initial results and provided us with insights for forthcoming directions.

In future work, we plan to develop an evaluation framework to account for the domain-expertise factor. We will also continue developing our guidance taxonomy and mechanisms to reflect better the lessons learned from the current work. Additionally, we intend to explore the potential of provenance techniques in informing uncertainty guidance. Another valuable future direction is the impact of explainable guidance on the user’s trust and confidence—especially in complex uncertainty scenarios. We anticipate that our approach (with modifications) can be generalized to other uncertainty visualizations applications outside the medical domain, such as weather forecasting, financial prediction, or material sciences.

CRedit authorship contribution statement

Maath Musleh: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review & editing, Visualization. **Ludvig Paul Muren:** Conceptualization, Validation, Resources, Data curation, Writing – review & editing. **Laura Toussaint:** Conceptualization, Validation, Resources, Data curation, Writing – review & editing. **Anne Vestergaard:** Conceptualization, Validation, Data curation. **Eduard Gröller:** Writing – review & editing. **Renata G. Raidou:** Conceptualization, Methodology, Validation, Formal analysis, Resources, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The data that has been used is confidential.

Acknowledgements

The authors would like to thank Dr. Davide Ceneda (TU Wien, Austria) for the fruitful discussions regarding guidance. Part of the research was enabled by VRVis, which is funded by BMK, BMDW, Styria, SFG, Tyrol, and Vienna Business Agency in the scope of COMET – Competence Centers for Excellent Technologies (879730), which is managed by FFG.

References

- [1] Ceneda D, Gschwandtner T, May T, Miksch S, Streit M, Tominski C. Guidance or no guidance? A decision tree can help. In: EuroVis workshop on visual analytics. The Eurographics Association; 2018, p. 19–23. <http://dx.doi.org/10.2312/eurova.20181107>.
- [2] Ceneda D, Gschwandtner T, Miksch S. A review of guidance approaches in visual data analysis: A multifocal perspective. *Comput Graph Forum* 2019;38:861–79. <http://dx.doi.org/10.1111/cgf.13730>.
- [3] Gillmann C, Smit NN, Gröller ME, Preim B, Vilanova A, Wischgoll T. Ten open challenges in medical visualization. *IEEE Comput Graph Appl* 2021;41(5):7–15. <http://dx.doi.org/10.1109/MCG.2021.3094858>.

- [4] Raidou RG. Uncertainty visualization: Recent developments and future challenges in prostate cancer radiotherapy planning. In: EuroVis workshop on reproducibility, verification, and validation in visualization. The Eurographics Association; 2018, p. 13–7. <http://dx.doi.org/10.2312/eurorv3.201811143>.
- [5] Saw CB, Li S, Battin F, McKeague J, Peters CA. External beam planning module of eclipse for external beam radiation therapy. *Med Dosim* 2018;43(2):195–204. <http://dx.doi.org/10.1016/j.meddos.2018.03.003>.
- [6] Schlachter M, Raidou RG, Muren LP, Preim B, Bühler K. State-of-the-art report: Visual computing in radiation therapy planning. *Comput Graph Forum* 2019;38(3):753–79. <http://dx.doi.org/10.1111/cgf.13726>.
- [7] Ristovski G, Preusser T, Hahn HK, Linsen L. Uncertainty in medical visualization: Towards a taxonomy. *Comput Graph* 2014;39:60–73. <http://dx.doi.org/10.1016/j.cag.2013.10.015>.
- [8] Gillmann C, Saur D, Wischgoll T, Scheuermann G. Uncertainty-aware visualization in medical imaging – A survey. *Comput Graph Forum* 2021;40(3):665–89. <http://dx.doi.org/10.1111/cgf.14333>.
- [9] Furmanová K, Grossmann N, Muren LP, Casares-Magaz O, Moiseenko V, Einck JP, et al. VAPOR: Visual analytics for the exploration of pelvic organ variability in radiotherapy. *Comput Graph* 2020;91:25–38. <http://dx.doi.org/10.1016/j.cag.2020.07.001>.
- [10] Furmanová K, Muren LP, Casares-Magaz O, Moiseenko V, Einck JP, Pilskog S, et al. PREVIS: Predictive visual analytics of anatomical variability for radiotherapy decision support. *Comput Graph* 2021;97:126–38. <http://dx.doi.org/10.1016/j.cag.2021.04.010>.
- [11] Suzuki J, Tateoka K, Shima K, Yaegashi Y, Fujimoto K, Saitoh Y, et al. Uncertainty in patient set-up margin analysis in radiation therapy. *J Radiat Res* 2012;53(4):615–9. <http://dx.doi.org/10.1093/jrr/rrs003>.
- [12] Paganetti H. Proton relative biological effectiveness - Uncertainties and opportunities. *Int J Part Therapy* 2018;5(1):2–14. <http://dx.doi.org/10.14338/IJPT-18-00011.1>.
- [13] Mohan R, Peeler C, Guan F, Bronk L, Cao W, Grosshans D. Radiobiological issues in proton therapy. *Acta Oncol* 2017;56(11):1367–73. <http://dx.doi.org/10.1080/0284186X.2017.1348621>.
- [14] Hofmaier J, Dedes G, Carlson D, Parodi K, Belka C, Kamp F. Variance-based sensitivity analysis for uncertainties in proton therapy: A framework to assess the effect of simultaneous uncertainties in range, positioning and RBE model predictions on RBE-weighted dose distributions. *Med Phys* 2020;48(2):805–18. <http://dx.doi.org/10.1002/mp.14596>.
- [15] Rørvik E, Fjæra LF, Dahle TJ, Dale JE, Engeseth GM, Stokkevåg CH, et al. Exploration and application of phenomenological RBE models for proton therapy. *Phys Med Biol* 2018;63(18):185013. <http://dx.doi.org/10.1088/1361-6560/aad9db>.
- [16] Kamp F, Wilkens J. Application of variance-based uncertainty and sensitivity analysis to biological modeling in carbon ion treatment plans. *Med Phys* 2019;46(2):437–47. <http://dx.doi.org/10.1002/mp.13306>.
- [17] Otterlei OM, Indelicato DJ, Toussaint L, Ytre-Hauge KS, Pilskog S, Fjæra LF, et al. Variation in relative biological effectiveness for cognitive structures in proton therapy of pediatric brain tumors. *Acta Oncol* 2021;60(2):267–74. <http://dx.doi.org/10.1080/0284186X.2020.1840626>.
- [18] Ceneda D, Gschwandtner T, May T, Miksch S, Schulz HJ, Streit M, et al. Amending the characterization of guidance in visual analytics. 2017. [arXiv:1710.06615](https://arxiv.org/abs/1710.06615).
- [19] Ceneda D, Andrienko N, Andrienko G, Gschwandtner T, Miksch S, Piccolotto N, et al. Guide me in analysis: A framework for guidance designers. *Comput Graph Forum* 2020;39(6):269–88. <http://dx.doi.org/10.1111/cgf.14017>.
- [20] Ceneda D, Gschwandtner T, May T, Miksch S, Schulz HJ, Streit M, et al. Characterizing guidance in visual analytics. *IEEE Trans Vis Comput Graph* 2017;23(1):111–20. <http://dx.doi.org/10.1109/TVCG.2016.2598468>.
- [21] Sperrle F, Jeitler A, Bernard J, Keim D, El-Assady M. Co-adaptive visual data analysis and guidance processes. *Comput Graph* 2021;100:93–105. <http://dx.doi.org/10.1016/j.cag.2021.06.016>.
- [22] Stoiber C, Ceneda D, Wagner M, Schetinger V, Gschwandtner T, Streit M, et al. Perspectives of visualization onboarding and guidance in VA. *Vis Inf* 2022;6(1):68–83. <http://dx.doi.org/10.1016/j.visinf.2022.02.005>.
- [23] Stoiber C, Walchshofer C, Grassinger F, Stitz H, Streit M, Aigner W. Design and comparative evaluation of visualization onboarding methods. In: The 14th international symposium on visual information communication and interaction. 2021, p. 1–5. <http://dx.doi.org/10.1145/3481549.3481558>.
- [24] Torsney-Weir T, Afrozeh S, Sedlmair M, Möller T. Risk fixers and sweet spotters: A study of the different approaches to using visual sensitivity analysis in an investment scenario. In: Proceedings of the Eurographics/IEEE VGTC conference on visualization: Short papers. Eurographics Association; 2018, p. 119–23. <http://dx.doi.org/10.2312/eurovisshort.20181089>.
- [25] Bögl M, Aigner W, Filzmoser P, Lammarsch T, Miksch S, Rind A. Visual analytics for model selection in time series analysis. *IEEE Trans Vis Comput Graph* 2013;19(12):2237–46. <http://dx.doi.org/10.1109/TVCG.2013.222>.
- [26] Torsney-Weir T, Sedlmair M, Möller T. Decision making in uncertainty visualization. In: VDMU workshop on visualization for decision making under uncertainty 2015. 2015, p. 1–4. <http://dx.doi.org/10.6084/M9.FIGSHARE.1585848>.
- [27] Schlachter M, Preim B, Bühler K, Raidou RG. Principles of visualization in radiation oncology. *Oncology* 2020;98(6):412–22. <http://dx.doi.org/10.1159/000504940>.
- [28] Budiarto E, Keijzer M, Storchi P, Hoogeman M, Bondar L, Mutanga T, et al. A population-based model to describe geometrical uncertainties in radiotherapy: Applied to prostate cases. *Phys Med Biol* 2011;56(4):1045–61. <http://dx.doi.org/10.1088/0031-9155/56/4/011>.
- [29] Berger W, Piringner H, Filzmoser P, Gröller E. Uncertainty-aware exploration of continuous parameter spaces using multivariate prediction. *Comput Graph Forum* 2011;30(3):911–20. <http://dx.doi.org/10.1111/j.1467-8659.2011.01940.x>.
- [30] Brodlie K, Allendes Osorio R, Lopes A. Expanding the frontiers of visual analytics and visualization. London: Springer London; 2012, p. 81–109. http://dx.doi.org/10.1007/978-1-4471-2804-5_6, chap. A Review of Uncertainty in Data Visualization.
- [31] Bonneau GP, Hege H-C, Johnson CR, Oliveira MM, Potter K, Rheingans P, et al. Overview and state-of-the-art of uncertainty visualization. In: Scientific visualization: Uncertainty, multifield, biomedical, and scalable visualization. London: Springer London; 2014, p. 5–30. http://dx.doi.org/10.1007/978-1-4471-6497-5_1, chap. 1.
- [32] Calvert S, Rypkema J, Holleman B, Azulay D, Jong A. Visualisation of uncertainty in probabilistic traffic models for policy and operations. *Transportation* 2017;44:701–29. <http://dx.doi.org/10.1007/s11116-015-9673-3>.
- [33] Belyakov S, Bozhenyuk A, Rozenberg I. Guidance in the visual analytics of cartographic images in the decision-making process. In: Advances in soft computing: 19th Mexican international conference on artificial intelligence, MICAI 2020, Mexico City, Mexico, October 12–17, 2020, Proceedings, Part I. Berlin, Heidelberg: Springer-Verlag; 2020, p. 351–69. http://dx.doi.org/10.1007/978-3-030-60884-2_26.
- [34] Floricel C, Nipu N, Biggs M, Wentzel A, Canahuate G, Dijk LV, et al. THALIS: Human-machine analysis of longitudinal symptoms in cancer therapy. *IEEE Trans Vis Comput Graph* 2022;28:151–61. <http://dx.doi.org/10.1109/TVCG.2021.3114810>.
- [35] Müller J, Stoehr M, Oeser A, Gaebel J, Streit M, Dietz A, et al. A visual approach to explainable computerized clinical decision support. *Comput Graph* 2020;91:1–11. <http://dx.doi.org/10.1016/j.cag.2020.06.004>.
- [36] Kamal A, Dhakal P, Javaid AY, Devabhaktuni VK, Kaur D, Zaiantz J, et al. Recent advances and challenges in uncertainty visualization: A survey. *J Vis* 2021;24(5):861–90. <http://dx.doi.org/10.1007/s12650-021-00755-1>.
- [37] Weissleder R, Wittenberg J, Harisinghani MG, Chen JW. Chapter 14: Imaging physics. In: Primer of diagnostic imaging. 5th ed. Philadelphia: Mosby; 2011, p. 690–746. <http://dx.doi.org/10.1016/B978-0-323-06538-2.00014-7>.
- [38] Luciani T, Burks A, Sugiyama C, Komperda J, Marai GE. Details-first, show context, overview last: Supporting exploration of viscous fingers in large-scale ensemble simulations. *IEEE Trans Vis Comput Graph* 2019;25(1):1225–35. <http://dx.doi.org/10.1109/TVCG.2018.2864849>.
- [39] Plumlee M, Ware C. Zooming versus multiple window interfaces: Cognitive costs of visual comparisons. *ACM Trans Comput-Hum Interact* 2006;13(2):179–209. <http://dx.doi.org/10.1145/1165734.1165736>.
- [40] Kehrler J, Piringner H, Berger W, Gröller ME. A model for structure-based comparison of many categories in small-multiple displays. *IEEE Trans Vis Comput Graph* 2013;19(12):2287–96. <http://dx.doi.org/10.1109/TVCG.2013.122>.
- [41] Tufte ER. *The visual display of quantitative information*. Cheshire, Connecticut: Graphics Press; 2001.
- [42] Wentzel A, Hanula P, Luciani T, Elgohari B, Elhalawani H, Canahuate G, et al. Cohort-based T-SSIM visual computing for radiation therapy prediction and exploration. *IEEE Trans Vis Comput Graph* 2019;26(1):949–59.
- [43] Hintze JL, Nelson RD. Violin plots: A box plot-density trace synergism. *Am Stat* 1998;52(2):181–4.
- [44] Zaghian M, Khabazian A, Lim G. A chance-constrained programming framework to handle uncertainties in radiation therapy treatment planning. *Eur J Oper Res* 2017;266(2):736–45. <http://dx.doi.org/10.1016/j.ejor.2017.10.018>.
- [45] Musleh M, Chatzimpampas A, Jusufi I. Visual analysis of blow molding machine multivariate time series data. *J Vis* 2022;25:1329–42. <http://dx.doi.org/10.1007/s12650-022-00857-4>.

- [46] Wall E, Agnihotri M, Matzen L, Divis K, Haass M, Endert A, et al. A heuristic approach to value-driven evaluation of visualizations. *IEEE Trans Vis Comput Graph* 2019;25(1):491–500. <http://dx.doi.org/10.1109/TVCG.2018.2865146>.
- [47] Brooke J. SUS–A quick and dirty usability scale. *Usability Eval Ind* 1996;189(194):4–7.
- [48] Lam H, Bertini E, Isenberg P, Plaisant C, Carpendale S. Empirical studies in information visualization: Seven scenarios. *IEEE Trans Vis Comput Graph* 2012;18(9):1520–36. <http://dx.doi.org/10.1109/TVCG.2011.279>.
- [49] Stitz H, Luger S, Streit M, Gehlenborg N. AVOCADO: Visualization of workflow-derived data provenance for reproducible biomedical research. *Comput Graph Forum* 2016;35(3):481–90. <http://dx.doi.org/10.1111/cgf.12924>.