In vivo and predictive interplay evaluation methodology for lung and esophageal cancer patients treated in free breathing with IMPT

Giorgio Cartechini[1]**, Esther Kneepkens**[1]**, Gloria Vilches-Freixas**[1]**, Indra Lubken**[1]**, Marije Velders**[1]**, Sebastiaan Nijsten**[1]**, Mirko Unipan**[1]**, and Ilaria Rinaldi**[1]

[1] Maastricht University Medical Centre+, Department of Radiation Oncology (Maastro), GROW School for Oncology and Reproduction, Maastricht, The Netherlands

Abstract.
Purpose: Pencil beam scanning proton therapy is sensitive to respiratory motion, leading to potential dose inhomogeneities due to interplay effects. We developed and validated a predictive framework to assess interplay and motion robustness in lung and esophageal cancer patients treated under free-breathing conditions.
Methods: A synthetic-breathing-based predictive model was implemented in the RayStation treatment planning system (TPS) and validated against an *in vivo* approach using patient-specific respiratory traces and machine log files. Both were benchmarked against the Monte Carlo engine FRED. To demonstrate the methodology, the framework was applied to two clinical cases treated on the Mevion S250i system without rescanning. Dose accumulation incorporated respiratory phase, range ($\pm3\%$), and setup (±5 mm) uncertainties.
Results: Excellent agreement was observed between TPS and FRED ($<1\%$ mean dose difference) and between predictive and *in vivo* models ($< 2\%$ across DVH metrics). Cumulative dose distributions for the primary CTV converged after five fractions, confirming robust delivery.
Conclusion: This automated, clinically integrated framework enables pre-treatment prediction and *in vivo* validation of interplay and motion robustness in PBS plans. Preliminary results support its clinical utility, especially for hypofractionation and targets with large motion (>2 cm). A larger cohort study is ongoing and will be reported separately.

Keywords: Proton therapy, interplay, motion robustness, in vivo dosimetry, Monte Carlo, lung cancer, esophageal cancer.

1. Introduction

Pencil beam scanning (PBS) proton therapy, particularly in the form of intensity-modulated proton therapy (IMPT), is the most advanced proton delivery technique, offering excellent dose conformity and enhanced sparing of surrounding healthy tissues [Chang et al., 2017]. However, its layer-by-layer, spot-scanning delivery is inherently sensitive to respiratory motion, particularly in thoracic and upper abdominal tumors [Chang et al., 2014]. This motion sensitivity can lead to the so-called interplay effect, a temporal mismatch between tumor motion and beam delivery, resulting in dose inhomogeneities such as cold and hot spots within the target volume [Bert and Durante, 2011].

Several studies have investigated the magnitude and clinical implications of this effect. Li et al. [Li et al., 2014] quantified interplay-induced deviations in stage III lung cancer patients, showing that although fractionation mitigates some heterogeneity, significant variations can still occur in individual fractions. Complementary, Monte Carlo simulations by Grassberger et al. [Grassberger et al., 2015] further demonstrated that small spot sizes and rapid energy switching can exacerbate dose non-uniformity, particularly in single-fraction deliveries.

To mitigate the impact of respiratory motion, several strategies have been investigated, including 4D robust optimization [Engwall et al., 2018, Spautz et al., 2023, Bengtsson et al., 2025] and various techniques, including rescanning [Rana and Rosenfeld, 2021, Knopf et al., 2011]. Despite their effectiveness, these methods can be complex and time-consuming, limiting their routine clinical implementation. As a result, more streamlined and time-efficient approaches are often preferred. At our institution, patients with lung [Taasti et al., 2025b] and esophageal cancers [Canters et al., 2024] are treated in free-breathing using 3D robust optimization based on an Internal Target Volume (ITV), defined as the union of Clinical Target Volume (CTV) contours across all respiratory phases. This ITV-based strategy, combined with planning on an average 4DCT and 3D and 4D robust evaluations [Taasti et al., 2021], incorporates motion into the optimization while preserving clinical feasibility.

Treatments are delivered using the Mevion Hyperscan S250i system (Mevion Medical Systems, Littleton, MA, USA) [Vilches-Freixas et al., 2020], which accelerates a fixed energy beam of around 230 MeV. The beam energy is modulated via eighteen range shifter plates within the nozzle. This configuration results in relatively large spot sizes, advantageous characteristics for mitigating the impact of motion in PBS delivery [Rana and Rosenfeld, 2022].

The so-called '4D evaluation' framework implemented at our institution to assess plan robustness [Taasti et al., 2021] accounts for anatomical variations across the breathing cycle but does not incorporate the temporal relationship between beam delivery and respiratory motion. As a result, it does not fully capture the dosimetric consequences of the interplay effect. To address this limitation, we propose a comprehensive framework for evaluating both interplay effects and motion robustness

in PBS plans for lung and esophageal cancer patients treated under free-breathing conditions.

We developed a predictive method to simulate interplay effects over the entire treatment course using synthetic respiratory scenarios, enabling evaluation to be performed prior to the start of treatment. The model is validated against an *in vivo* approach that integrates fraction-specific respiratory traces and machine log files recorded during treatment. The evaluation framework is fully automated, implemented within the clinical treatment planning system, and further validated using the independent Monte Carlo dose engine FRED [Schiavi et al., 2017], clinically validated in our proton facility [Gajewski et al., 2020]. This article focuses on presenting the methodology for both predictive and *in vivo* evaluations; extension of this work to a larger patient cohort is currently underway and will be the subject of a separate publication.

2. Materials and Methods

2.1. Interplay effect evaluation: Predictive and in vivo methods

The interplay evaluation was conducted using two approaches: predictive and *in vivo*. The predictive approach is designed to be performed before the start of the treatment, without prior knowledge of the patient's breathing trace. In contrast, the *in vivo* approach retrospectively evaluates plan robustness to organ motion by incorporating the patient-specific breathing traces per fraction acquired during treatment and the corresponding machine log files recorded after each fraction.

In vivo and predictive interplay evaluations were performed using two Monte Carlo dose engines: FRED v3.76.4 [Schiavi et al., 2017], and the clinical treatment planning system (TPS) RayStation (RaySearch Laboratories, Stockholm, Sweden) version 12B (Monte Carlo v5.4). The interplay workflow is illustrated in Figure 1. The 4DCT, breathing traces, and a machine log file are used as input for both Monte Carlo systems. The 4DCT represents the anatomical model of the patient's geometry. It consists of eight 3D CTs corresponding to different respiratory phases: 0%, 25%, 50%, 75%, and 100% inhale, and 25%, 50%, 75% exhale. The ANZAI belt system (Anzai Medical Co. Ltd, Tokyo, Japan) was used to measure the patient's respiratory amplitude during the CT scan. The 3D volumetric images were sorted based on the acquired breathing signals to assign them to the appropriate respiratory phases [Lu et al., 2006, Li et al., 2006].

The *in vivo* approach exploits the fraction-specific breathing traces acquired by the C-RAD system [Walter et al., 2016] installed in the treatment room. Since the predictive approach is intended to be performed before treatment is delivered and patient-specific breathing traces are not available, we generated 24 synthetic breathing curves based on a sinusoidal function model, varying the breathing period (2 s, 3.5 s, and 5 s) and the starting phase according to the eight CT phases. By comparing predictive and *in vivo* approaches, we will show that these 24 traces represent a range of plausible

breathing scenarios for the patient.

To map each beam spot to one of the eight phases of the 4DCT, we used fraction-specific machine log files for the *in vivo* model. For the predictive model, log files can be generated before the start of treatment through a dry run of the plan on a water phantom, similar to the patient quality assurance (QA) procedure.

Following the methodology described by Pastor-Serrano et al. [Pastor-Serrano et al., 2021], we combined 4DCT phases, patient-specific breathing traces, and machine log file data to assign each proton spot to a corresponding 4DCT phase. Dose was then recalculated with both FRED and TPS Monte Carlo on all eight 4DCT phases. Using RayStation's deformable image registration (DIR) algorithm, the resulting dose distributions were warped onto a reference phase (50% exhale) to compute the cumulative dose (Figure 1).

To assess the robustness of the plan against organ motion, as well as setup and range uncertainties, we systematically applied an isocenter shift of ±5 mm along each Cartesian axis, combined with a systematic range error of ±3% in the CT calibration curve. To reduce computational time, we employed a probabilistic approach: for each breathing scenario, one random combination of range and setup error (representing a worst-case scenario) was sampled.

The effect of fractionation was quantified by accumulating the dose distribution over the entire treatment course, as illustrated in Figure 2. For the predictive method, a synthetic treatment course was simulated by randomly assigning one of the 24 breathing scenarios to each treatment fraction. The cumulative dose across all fractions was then calculated, and dose–volume histogram (DVH) parameters were extracted. Specifically, for each cumulative dose distribution, DVH metrics were computed for the primary tumor CTV (CTVp) and nodal CTV (CTVn), including D98 and D95 (minimum dose received by 98% and 95% of the volume, respectively) and V95 (volume receiving at least 95% of the prescribed dose). To assess variability, the simulation was repeated ten times, and the mean and standard deviation of the DVH parameters are reported.

For the *in vivo* evaluation, we recalculated the dose distribution from the fraction-specific breathing trace and machine log file, and then accumulated the doses over all fractions. As described previously, we extracted the DVH parameters D98, D95, and V95.

To quantitatively compare the inter-fractional distribution of DVH parameters obtained from the *in vivo* evaluation and the predicted scenarios, we applied the statistical analysis method described by Pastor-Serrano et al. [Pastor-Serrano et al., 2021]. For each quantity of interest Θ (i.e., D98, D95 and V95), we approximated the distribution across the 24 breathing trace scenarios (or treatment fraction for the *in vivo*) using a percentile vector, $\delta_\Theta = \{median, 2, 98\}$. We then computed the Relative Difference Error (RDE), defined as the average relative difference between two percentile vectors associated with different distributions of the same quantity of interest:

$$RDE(\delta_{\Theta,1}, \delta_{\Theta,2}) = \frac{1}{n_i} \sum_i^{n_i} \frac{|\Theta_{i,1} - \Theta_{i,2}|}{\Theta_{i,2}} \times 100 \qquad (1)$$

where n_i is the number of elements in the percentile vector.

2.2. FRED: Independent Monte Carlo engine

Dose distributions were recalculated using the fast GPU-based Monte Carlo dose engine FRED [Schiavi et al., 2017], accounting for both organ motion and the temporal dynamics of beam delivery. The Mevion S250i beam model was implemented in FRED and clinically validated against experimental commissioning data to accurately reproduce the beam characteristics used at our institution [Gajewski et al., 2020].

The 4DCT was imported into FRED by loading the eight 3D CT phases as standard geometry volumes. Efficient RAM and GPU memory management were required due to the large memory footprint of storing all CT phases simultaneously.

FRED supports dynamic activation and deactivation of geometry components during simulation. Leveraging this functionality, we activated the corresponding 4DCT phase during the delivery of each individual spot to model motion-resolved dose deposition. To account for setup and range uncertainties, we introduced perturbations in each simulation by modifying the Hounsfield Unit–Relative Stopping Power (HU-RSP) calibration curve by $\pm 3\%$ and by applying isocenter shifts of ± 5 mm in all directions. As a result, eight separate dose distributions were generated, each corresponding to one of the 4DCT phases.

The resulting dose distributions were imported into the RayStation TPS, where deformable image registration was applied to map each distribution onto a common reference phase. The final accumulated dose was then computed by summing the warped dose distributions.

2.3. RayStation TPS Evaluation

RayStation version 12B was used for dose recalculation, employing the Monte Carlo v5.4 algorithm with a final dose calculation uncertainty of 1%. To streamline the workflow, we integrated the interplay evaluation directly into the clinical treatment planning system using a Python (v3.8) script via the RayStation API, automating the steps outlined in Figure 1. The evaluation followed the same principles as those implemented in FRED: Control points were mapped to the corresponding 4DCT phases, and dose distributions were recalculated for each phase. Range and setup uncertainties were introduced by perturbing the dose distributions accordingly. As in the FRED-based approach, RayStation's DIR algorithm was used to map all eight phase-specific dose distributions onto a common reference phase, after which the cumulative dose was computed.

2.4. Lung and Esophageal Cancer Patient Cases

We applied the proposed interplay evaluation methodology—comprising both predictive and *in vivo* approaches—to two clinical cases, a lung and an esophageal cancer patient treated at our institution to test the methodology workflows and assess clinical feasibility (Figure 3).
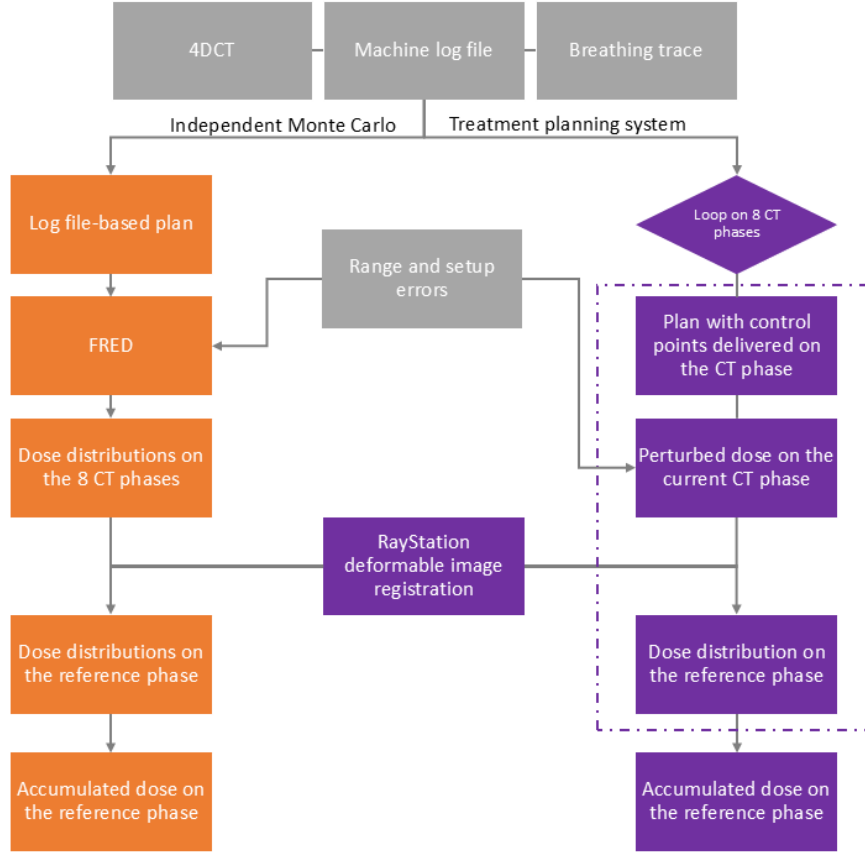
Figure 1: Workflow for the interplay evaluation using the FRED independent Monte Carlo engine (orange blocks) and the clinical TPS, RayStation (purple blocks). The dashed line groups the blocks within the loop over the 8 4DCT phases. The grey blocks represent the general input models independently of the dose engine used. This workflow was applied to both the predictive and *in vivo* approaches.

The lung cancer patient was characterized by a gross tumor volume (GTV) amplitude, defined as the maximum displacement of the GTV center of mass across 4DCT phases, of 9.7 mm. The treatment plan employed three posterior beams and was 3D robustly optimized on the ITV expanded by 1 mm to account for delineation uncertainties, incorporating 3% range and 5 mm setup errors. Further details on the lung cancer patient's plan setup and optimization are provided by Taasti et al.[Taasti et al., 2025b].

Similar to the lung case, the esophageal cancer patient was planned using three posterior beams and robustly optimized on the ITV expanded by 3 mm, incorporating 3% range and 5 mm setup uncertainties. Additional details are provided by Canters et al. [Canters et al., 2024].
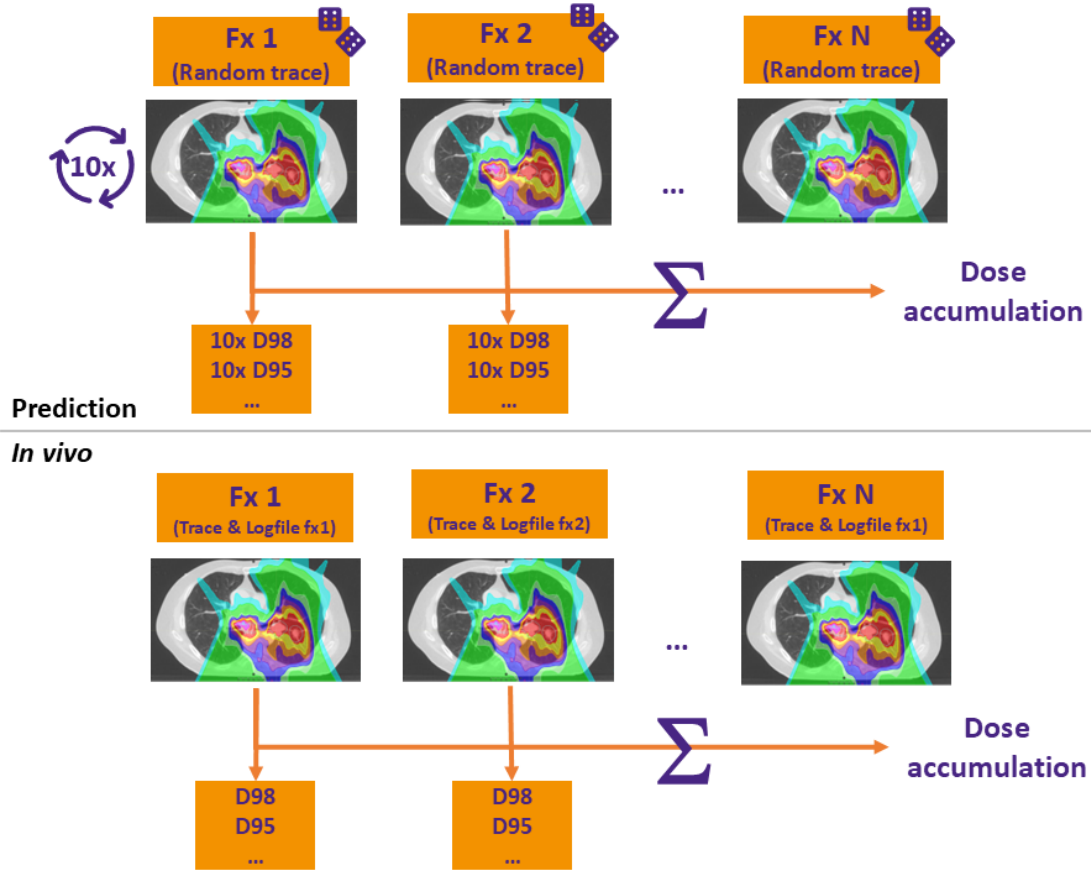
Figure 2: Top: Dose accumulation over treatment fractions for the predictive approach. A synthetic fractionation scheme was generated by randomly sampling a dose distribution scenario for each treatment fraction. This process was repeated ten times to assess variability. Bottom: Dose accumulation for the *in vivo* approach using fraction-specific log files and breathing traces for each treatment session.

3. Results

This section presents a systematic comparison between (i) the FRED Monte Carlo dose engine and the RayStation TPS for interplay evaluation, and (ii) the predictive and *in vivo* methods for modeling respiratory motion. The comparison between FRED and RayStation TPS was conducted using a lung cancer case, while the analysis of predictive versus *in vivo* approaches was performed for both lung and esophageal cancer patients. Unless otherwise stated, all results reported in Section 3.2 were generated using the FRED dose engine.

3.1. Comparison between FRED and RayStation TPS

To verify consistency between dose engines, we compared interplay-affected dose distributions computed with FRED and RayStation using the *in vivo* method. As
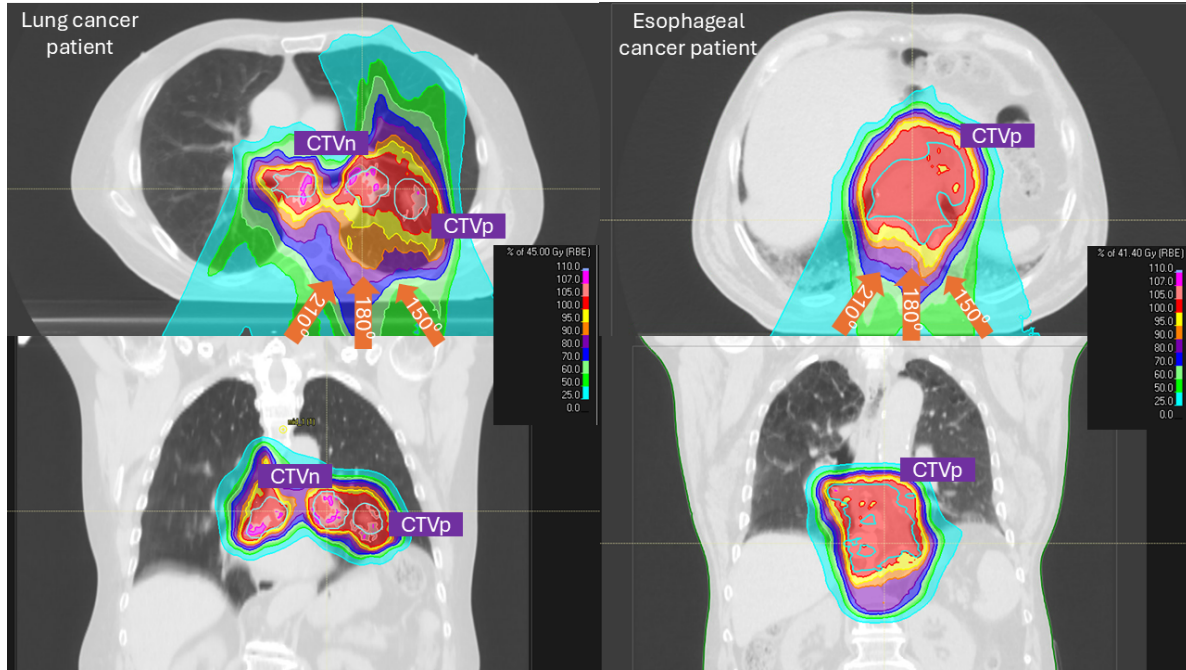
Figure 3: Axial and coronal view of the lung and esophageal cancer patients. Both cases were irradiated with three beams at 150°, 180° and 210° gantry angles.

illustrated in Figure 4, both engines yielded similar DVH parameter values for both the CTVp and CTVn. The median absolute difference between the two dose engines was below 1% for all metrics, with only a single fraction showing a deviation above 2% for D98. RDE values for D98, D95, and V95 remained consistently below 1%, confirming a high level of agreement between the two dose calculation engines.

3.2. Interplay Evaluation: Predictive vs. In Vivo Approach

The predictive method, based on 24 synthetic sinusoidal breathing scenarios, was compared against the *in vivo* model that integrates patient-specific breathing traces and machine log files. Figure 5 shows the fractional DVH distributions (D98, D95, and V95) for both lung (panel a) and esophageal (panel b) cancer patients. The violin plots demonstrate that the predictive method accurately reproduces the dosimetric variability observed in the *in vivo* calculations. Quantitatively, the RDE between the two approaches was below 1% for both CTVp and CTVn, confirming strong concordance.

To investigate the mitigation of interplay effects through fractionation, we analyzed cumulative DVH metrics across the treatment course (Figure 6). In these plots, the predictive model is represented by a shaded band (indicating standard deviation from 10 random fractionation schemes) and a dashed line (mean value). The *in vivo* results are shown as orange circles.

In the lung cancer case (Figure 6(a)), the cumulative DVH metrics for the primary tumor (CTVp) converged to within 2% of the nominal plan after just five fractions,
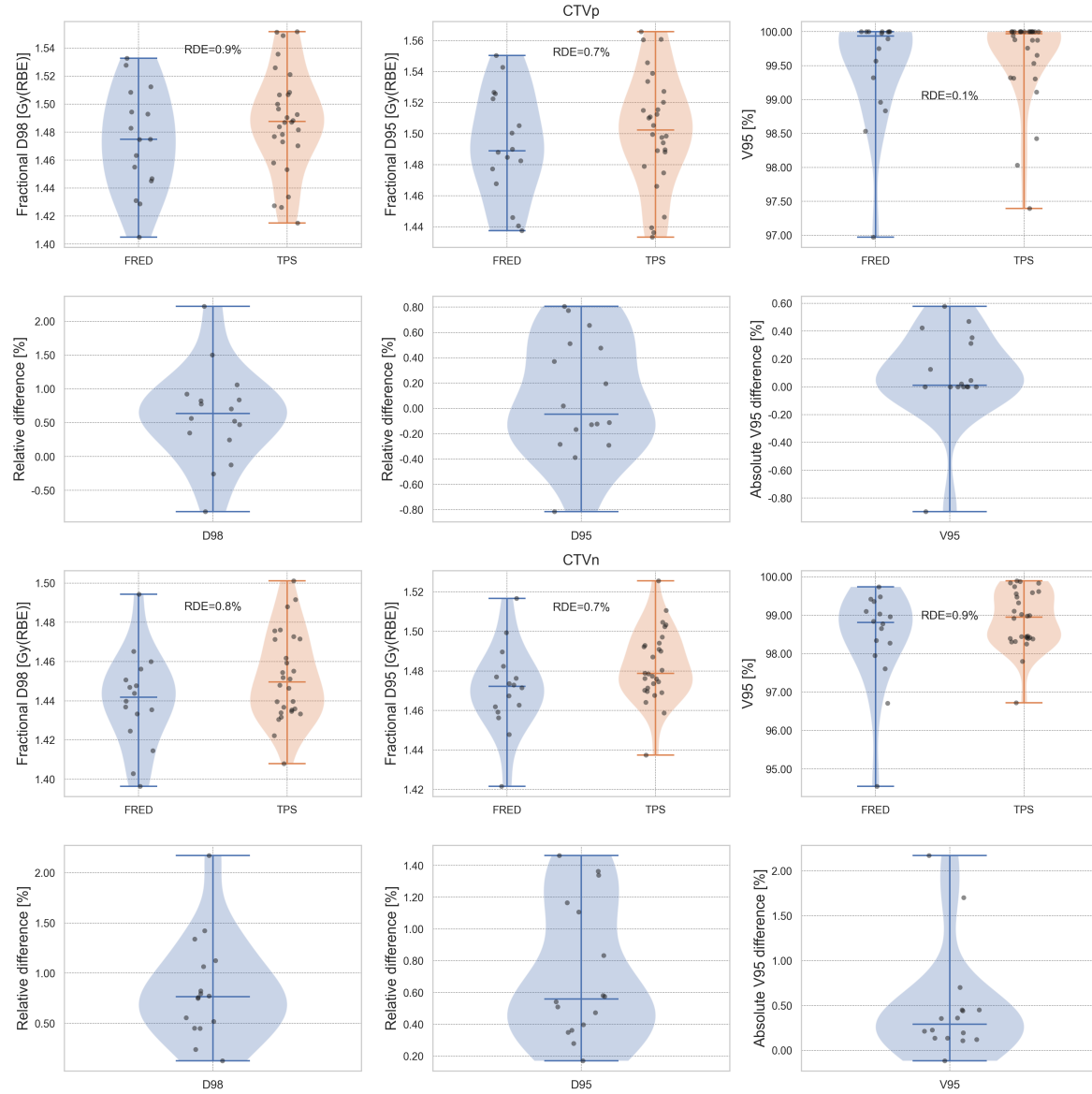
Figure 4: Comparison of D98, D95, and V95 between RayStation TPS and FRED dose engines using the *in vivo* approach for the lung cancer case.

indicating that the interplay effect was effectively averaged out early in the treatment course. A similar trend was observed for the nodal volume (CTVn), although the cumulative dose remained slightly lower than planned throughout the treatment. At the end of the course, the D98 and D95 for CTVn were $2.0\% \pm 0.2\%$ and $1.3\% \pm 0.2\%$ below the nominal plan, respectively. It is important to note that, following our current clinical practice, all nodal regions are treated as a single structure (CTVn), and no motion management or volume expansion—such as an internal target volume (ITV) is applied to CTVn [Taasti et al., 2025b].

In the esophageal cancer case (Figure 6(b)), both the predictive and *in vivo* methods produced cumulative DVH values that remained within 2% of the planned dose starting
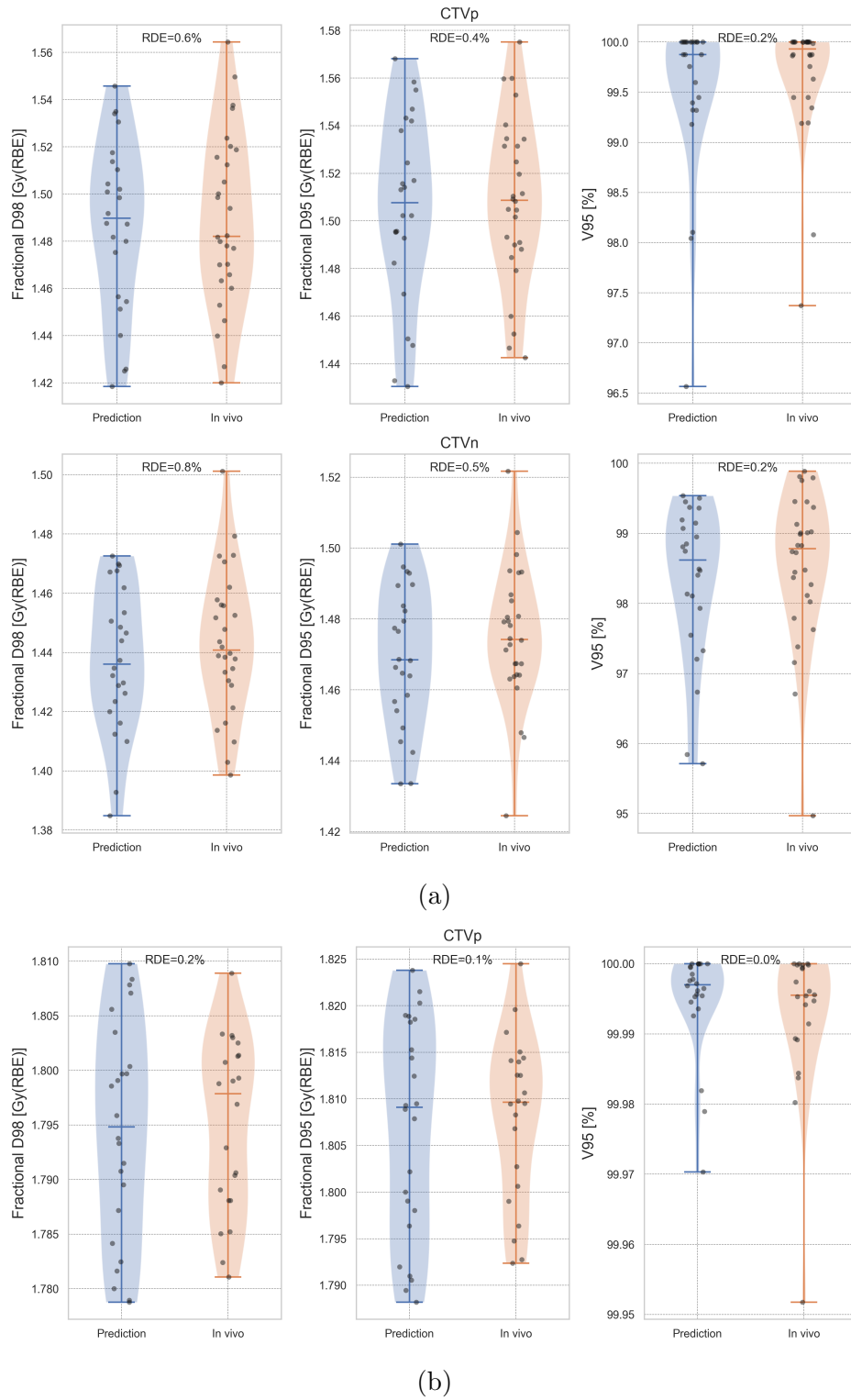
(a)

(b)

Figure 5: Violin plots of the D98, D95, and V95 distributions for the predictive and *in vivo* methods computed with FRED Monte Carlo dose engine. (a) Lung cancer patients (CTVp and CTVn); (b) Esophageal cancer patients (CTVp only).

from the first fraction. This close agreement across all fractions highlights the robustness of the treatment plan under free-breathing conditions, even in the presence of respiratory motion.

4. Discussion

This study presents a comprehensive framework for predicting interplay effects combined with range and setup uncertainties for a Mevion Hyperscan S250i PBS proton therapy system. The focus is on lung and esophageal cancer patients treated under free-breathing conditions. We proposed a predictive interplay assessment method based on synthetic breathing traces and compared it to an *in vivo* approach that incorporates patient-specific respiratory traces and machine log files. Both methods were implemented within the RayStation treatment planning system and validated using the independent Monte Carlo dose engine FRED.

The two dose engines demonstrated excellent agreement, with mean dose differences below 1%, which is on the same order as the calculation uncertainty of the clinical TPS, thereby confirming the robustness of the implementation (Figure 4). Thanks to its high computational speed and exceptional versatility, FRED provides a powerful platform for advanced research and rapid development of novel interplay evaluation workflows. In parallel, the TPS—with its integrated graphical user interface—remains the more user-friendly environment for routine clinical use.

Application of the two approaches to clinical cases yielded highly consistent results, with relative differences below 2% across all evaluated DVH metrics (Figure 5). Although the predictive model assumes a sinusoidal breathing pattern and omits inter-fractional delivery variations, its strong agreement with the *in vivo* evaluation suggests that these simplifications have minimal impact on clinical interplay estimation [Pastor-Serrano et al., 2021]. Furthermore, the 24 breathing scenarios used in the predictive model effectively captured the range of *in vivo* dose variations.

The convergence of cumulative dose distributions toward the nominal plan (Figure 6) further confirms the well-established mitigation of interplay effects through fractionation [Seco et al., 2009, Grassberger et al., 2015]. For the primary CTV (CTVp), only minor deviations from the nominal dose were observed, even when accounting for respiratory motion, range uncertainties, and setup deviations. DVH parameters converged to within 2% of planned values after just five fractions, demonstrating plan robustness under conventional fractionation. These findings align with prior work by Meijers et al. [Meijers et al., 2020], who also found that organ motion has a relatively minor impact on dose degradation compared to anatomical changes—though their study employed rescanning.

Our results suggest that the relatively large spot sizes of the Mevion Hyperscan S250i system inherently mitigate interplay effects, eliminating the need for time-consuming rescanning techniques. This observation is consistent with previous works [Rana and Rosenfeld, 2022, Dowdell et al., 2013, Grewal et al., 2023] and has been
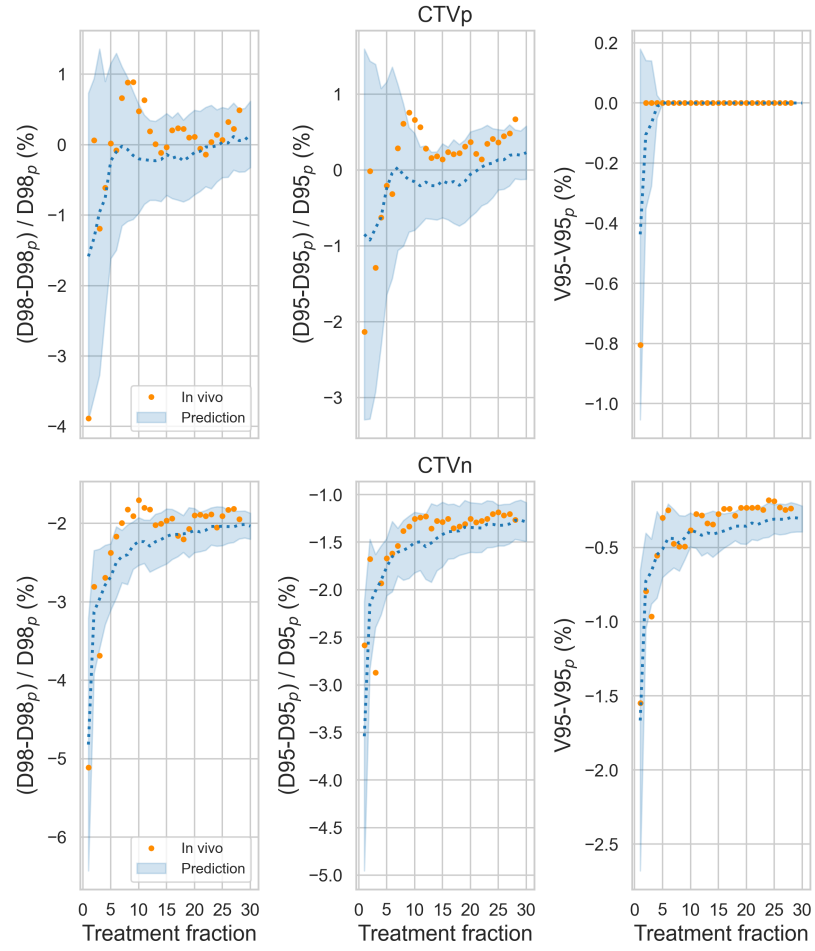
(a)
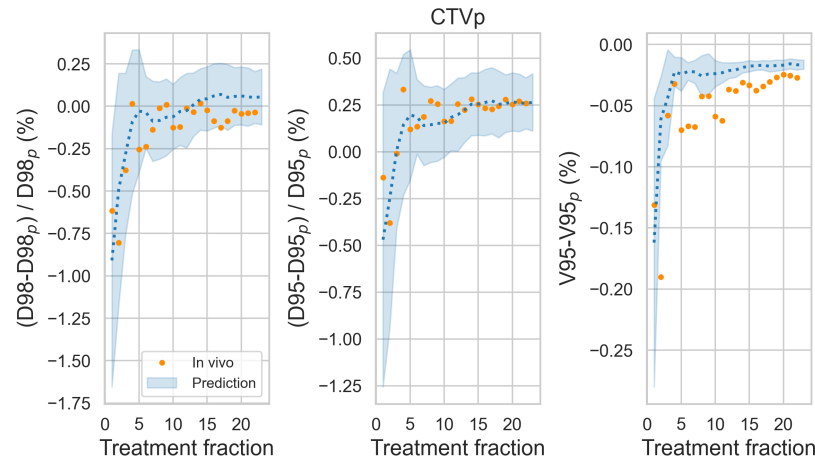
(b)

Figure 6: Cumulative DVH parameters across the treatment course for the predictive and *in vivo* approaches. (a) Lung cancer patient (CTVp and CTVn); (b) Esophageal cancer patient (CTVp only).

further substantiated by phantom studies conducted at our institution.

In the lung cancer case, a comparative analysis between the CTVp and CTVn revealed greater robustness for the former. The CTVp was optimized using 3D robust optimization based on the ITV, further expanded by 1 mm, whereas the CTVn was optimized without any additional margins, resulting in reduced dosimetric robustness. This confirms that the ITV robust optimization improves the robustness also towards organ motion and interplay [Inoue et al., 2016]. Despite some loss of coverage in the CTVp during the first five fractions, and a more pronounced degradation in the CTVn over the full treatment course, the biological implications may be limited, given the small magnitude of these deviations [Bortfeld and Paganetti, 2006].

Limitations

A primary limitation of this study is the exclusion of inter-fractional anatomical changes and variations in tumor motion amplitude throughout the treatment course. Interplay effects were assessed using the planning 4DCT for each fraction, which may limit generalization, particularly in patients experiencing significant anatomical evolution or respiratory variability. While DIR enables dose accumulation across repeated CTs, it introduces uncertainties, particularly in the presence of substantial geometric changes. Several studies, including Ribeiro et al. [Ribeiro et al., 2018] and Stützer et al. [Stützer et al., 2016], have highlighted both the capabilities and limitations of DIR in proton therapy. Before integrating accumulated dose evaluations into the predictive framework, DIR errors must be rigorously quantified. A recent study by Taasti et al. [Taasti et al., 2025a] introduced a method for dose accumulation that accounts for setup and range uncertainties. This approach considers anatomical changes between fractions and also facilitates the QA of deformable dose mapping.

On the other hand, data extracted from our lung patient cohort indicate a general trend of decreasing tumor motion amplitude over the course of treatment, supporting the assumption that the planning 4DCT remains a reasonable surrogate for respiratory motion modeling. Figure 7 shows the GTV amplitude of 24 patients treated at our institution from mid-2024 onward, computed as the maximum 1D displacement of the GTV center of mass over the three orthogonal directions. This trend was also confirmed by a previous study analyzing patients treated from 2019 to 2023 [Taasti et al., 2025b]. In cases requiring adaptive planning due to anatomical or motion changes, the predictive model can be reapplied using the updated treatment plan and planning 4DCT, which intrinsically accounts for the new tumor amplitude. Notably, the predictive evaluation described here is intended for pre-treatment use, when no information on future anatomical changes is available.

Another limitation relates to the *in vivo* model's assumption of phase alignment between the patient-specific breathing trace and the 4DCT phase at the start of treatment. This simplification can introduce uncertainties in interplay calculations, as shown in previous studies [Dowdell et al., 2013]. However, the predictive model
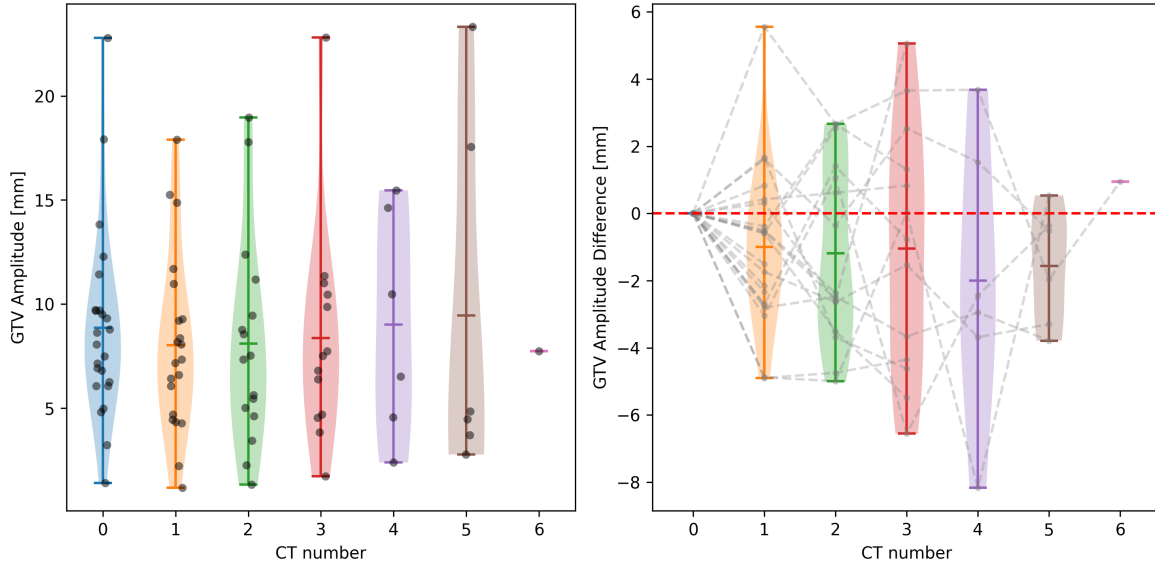
Figure 7: Left panel: distribution of the GTV amplitude as a function of repeated 4DCT number. The right panel shows the difference between the amplitude in the repeated CT and the first planning CT (CT number 0).

incorporates all initial phase scenarios across 4DCT phases, inherently addressing this uncertainty. The strong agreement between the predictive and *in vivo* models supports the validity of this assumption in practice.

5. Conclusions

In this study, we presented a predictive approach to evaluate the robustness of proton therapy plans against organ motion, range, and setup uncertainties before the start of treatment. The model was validated using *in vivo* data incorporating patient- and fraction-specific parameters, such as breathing signals and machine log files. Two clinical cases, one lung and one esophageal cancer patient treated under free-breathing conditions without rescanning, demonstrated only minor dose degradation in the primary CTV due to organ motion.

    While these results are encouraging, further investigation in a larger patient cohort is necessary to confirm these preliminary findings. This work is currently ongoing and will be presented in a separate publication. If validated, the proposed predictive framework could support clinical decision-making and enable the adoption of hypofractionated treatment regimens, as well as the treatment of lung targets exhibiting large respiratory motion (i.e., > 2cm).

Acknowledgments

References

[Bengtsson et al., 2025] Bengtsson, I., Forsgren, A., Fredriksson, A., and Zhang, Y. (2025). Interplay-robust optimization for treating irregularly breathing lung patients with pencil beam scanning. *Medical Physics*.

[Bert and Durante, 2011] Bert, C. and Durante, M. (2011). Motion in radiotherapy: particle therapy. *Physics in Medicine & Biology*, 56(16):R113.

[Bortfeld and Paganetti, 2006] Bortfeld, T. and Paganetti, H. (2006). The biologic relevance of daily dose variations in adaptive treatment planning. *International Journal of Radiation Oncology* Biology* Physics*, 65(3):899–906.

[Canters et al., 2024] Canters, R., van der Klugt, K., Taasti, V. T., Buijsen, J., Ta, B., Steenbakkers, I., Houben, R., Vilches-Freixas, G., and Berbee, M. (2024). Robustness of intensity modulated proton treatment of esophageal cancer for anatomical changes and breathing motion. *Radiotherapy and Oncology*, 198:110409.

[Chang et al., 2014] Chang, J. Y., Li, H., Zhu, X. R., Liao, Z., Zhao, L., Liu, A., Li, Y., Sahoo, N., Poenisch, F., Gomez, D. R., et al. (2014). Clinical implementation of intensity modulated proton therapy for thoracic malignancies. *International Journal of Radiation Oncology* Biology* Physics*, 90(4):809–818.

[Chang et al., 2017] Chang, J. Y., Zhang, X., Knopf, A., Li, H., Mori, S., Dong, L., Lu, H.-M., Liu, W., Badiyan, S. N., Both, S., et al. (2017). Consensus guidelines for implementing pencil-beam scanning proton therapy for thoracic malignancies on behalf of the ptcog thoracic and lymphoma subcommittee. *International Journal of Radiation Oncology* Biology* Physics*, 99(1):41–50.

[Dowdell et al., 2013] Dowdell, S., Grassberger, C., Sharp, G., and Paganetti, H. (2013). Interplay effects in proton scanning for lung: a 4d monte carlo study assessing the impact of tumor and beam delivery parameters. *Physics in Medicine & Biology*, 58(12):4137.

[Engwall et al., 2018] Engwall, E., Fredriksson, A., and Glimelius, L. (2018). 4d robust optimization including uncertainties in time structures can reduce the interplay effect in proton pencil beam scanning radiation therapy. *Medical physics*, 45(9):4020–4029.

[Gajewski et al., 2020] Gajewski, J., Schiavi, A., Krah, N., Vilches-Freixas, G., Rucinski, A., Patera, V., and Rinaldi, I. (2020). Implementation of a compact spot-scanning proton therapy system in a gpu monte carlo code to support clinical routine. *Frontiers in Physics*, 8:578605.

[Grassberger et al., 2015] Grassberger, C., Dowdell, S., Sharp, G., and Paganetti, H. (2015). Motion mitigation for lung cancer patients treated with active scanning proton therapy. *Medical physics*, 42(5):2462–2469.

[Grewal et al., 2023] Grewal, H. S., Ahmad, S., and Jin, H. (2023). Dosimetric study of the interplay effect using three-dimensional motion phantom in proton pencil beam scanning treatment of moving thoracic tumours. *Journal of Radiotherapy in Practice*, 22:e11.

[Inoue et al., 2016] Inoue, T., Widder, J., van Dijk, L. V., Takegawa, H., Koizumi, M., Takashina, M., Usui, K., Kurokawa, C., Sugimoto, S., Saito, A. I., et al. (2016). Limited impact of setup and range uncertainties, breathing motion, and interplay effects in robustly optimized intensity modulated proton therapy for stage iii non-small cell lung cancer. *International Journal of Radiation Oncology* Biology* Physics*, 96(3):661–669.

[Knopf et al., 2011] Knopf, A.-C., Hong, T. S., and Lomax, A. (2011). Scanned proton radiotherapy for mobile targets—the effectiveness of re-scanning in the context of different treatment planning approaches and for different motion characteristics. *Physics in Medicine & Biology*, 56(22):7257.

[Li et al., 2006] Li, X. A., Stepaniak, C., and Gore, E. (2006). Technical and dosimetric aspects of respiratory gating using a pressure-sensor motion monitoring system. *Medical physics*, 33(1):145–154.

[Li et al., 2014] Li, Y., Kardar, L., Li, X., Li, H., Cao, W., Chang, J. Y., Liao, L., Zhu, R. X., Sahoo,

N., Gillin, M., et al. (2014). On the interplay effects with proton scanning beams in stage iii lung cancer. *Medical physics*, 41(2):021721.

[Lu et al., 2006] Lu, W., Parikh, P. J., Hubenschmidt, J. P., Bradley, J. D., and Low, D. A. (2006). A comparison between amplitude sorting and phase-angle sorting using external respiratory measurement for 4d ct. *Medical physics*, 33(8):2964–2974.

[Meijers et al., 2020] Meijers, A., Knopf, A.-C., Crijns, A. P., Ubbels, J. F., Niezink, A. G., Langendijk, J. A., Wijsman, R., and Both, S. (2020). Evaluation of interplay and organ motion effects by means of 4d dose reconstruction and accumulation. *Radiotherapy and Oncology*, 150:268–274.

[Pastor-Serrano et al., 2021] Pastor-Serrano, O., Habraken, S., Lathouwers, D., Hoogeman, M., Schaart, D., and Perkó, Z. (2021). How should we model and evaluate breathing interplay effects in impt? *Physics in Medicine & Biology*, 66(23):235003.

[Rana and Rosenfeld, 2021] Rana, S. and Rosenfeld, A. B. (2021). Investigating volumetric repainting to mitigate interplay effect on 4d robustly optimized lung cancer plans in pencil beam scanning proton therapy. *Journal of Applied Clinical Medical Physics*, 22(3):107–118.

[Rana and Rosenfeld, 2022] Rana, S. and Rosenfeld, A. B. (2022). Small spot size versus large spot size: Effect on plan quality for lung cancer in pencil beam scanning proton therapy. *Journal of Applied Clinical Medical Physics*, 23(2):e13512.

[Ribeiro et al., 2018] Ribeiro, C. O., Knopf, A., Langendijk, J. A., Weber, D. C., Lomax, A. J., and Zhang, Y. (2018). Assessment of dosimetric errors induced by deformable image registration methods in 4d pencil beam scanned proton treatment planning for liver tumours. *Radiotherapy and Oncology*, 128(1):174–181.

[Schiavi et al., 2017] Schiavi, A., Senzacqua, M., Pioli, S., Mairani, A., Magro, G., Molinelli, S., Ciocca, M., Battistoni, G., and Patera, V. (2017). Fred: a gpu-accelerated fast-monte carlo code for rapid treatment plan recalculation in ion beam therapy. *Physics in Medicine & Biology*, 62(18):7482.

[Seco et al., 2009] Seco, J., Robertson, D., Trofimov, A., and Paganetti, H. (2009). Breathing interplay effects during proton beam scanning: simulation and statistical analysis. *Physics in Medicine & Biology*, 54(14):N283.

[Spautz et al., 2023] Spautz, S., Haase, L., Tschiche, M., Makocki, S., Richter, C., Troost, E. G., and Stützer, K. (2023). Comparison of 3d and 4d robustly optimized proton treatment plans for non-small cell lung cancer patients with tumour motion amplitudes larger than 5 mm. *Physics and Imaging in Radiation Oncology*, 27:100465.

[Stützer et al., 2016] Stützer, K., Haase, R., Lohaus, F., Barczyk, S., Exner, F., Löck, S., Rühaak, J., Lassen-Schmidt, B., Corr, D., and Richter, C. (2016). Evaluation of a deformable registration algorithm for subsequent lung computed tomography imaging during radiochemotherapy. *Medical physics*, 43(9):5028–5039.

[Taasti et al., 2025a] Taasti, V. T., Hattu, D., Berbée, M., de Ruysscher, D., and Canters, R. (2025a). Dose accumulation over the treatment course including setup and range uncertainty for proton therapy of esophageal and lung cancer patients. *Medical Physics*, 52(8):e18037.

[Taasti et al., 2021] Taasti, V. T., Hattu, D., Vaassen, F., Canters, R., Velders, M., Mannens, J., van Loon, J., Rinaldi, I., Unipan, M., and van Elmpt, W. (2021). Treatment planning and 4d robust evaluation strategy for proton therapy of lung tumors with large motion amplitude. *Medical Physics*, 48(8):4425–4437.

[Taasti et al., 2025b] Taasti, V. T., Kneepkens, E., van der Stoep, J., Velders, M., Cobben, M., Vullings, A., Buck, J., Visser, F., van den Bosch, M., Hattu, D., et al. (2025b). Proton therapy of lung cancer patients–treatment strategies and clinical experience from a medical physicist's perspective. *Physica Medica*, 130:104890.

[Vilches-Freixas et al., 2020] Vilches-Freixas, G., Unipan, M., Rinaldi, I., Martens, J., Roijen, E., Almeida, I. P., Decabooter, E., and Bosmans, G. (2020). Beam commissioning of the first compact proton therapy system with spot scanning and dynamic field collimation. *The British journal of radiology*, 93(1107):20190598.

[Walter et al., 2016] Walter, F., Freislederer, P., Belka, C., Heinz, C., Söhn, M., and Roeder, F. (2016).

Evaluation of daily patient positioning for radiotherapy with a commercial 3d surface-imaging system (catalyst™). *Radiation oncology*, 11(1):154.