
UNIVERSIDADE D
COIMBRA

Tânia Filipa Sobrinho dos Santos

# IMRT NATIONAL AUDIT – APPLICATION OF IAEA METHODOLOGY AND EXTENSION TO NEW TECHNOLOGIES

**Tese no âmbito do Doutoramento em Engenharia Física, Metrologia e Qualidade, orientada pela Professora Doutora Maria do Carmo Carrilho Calado Antunes Lopes, co-orientada pelo Professor Doutor Paulo Alexandre Vieira Crespo e apresentada ao Departamento de Física da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.**

Dezembro de 2020


UNIVERSIDADE D
COIMBRA

Tânia Filipa Sobrinho dos Santos

# IMRT national audit – application of IAEA methodology and extension to new technologies

Tese no âmbito do Doutoramento em Engenharia Física, Metrologia e Qualidade, orientada pela Professora Doutora Maria do Carmo Carrilho Calado Antunes Lopes, co-orientada pelo Professor Doutor Paulo Alexandre Vieira Crespo e apresentada ao Departamento de Física da Faculdade de Ciências e Tecnologia da Universidade de Coimbra.

Dezembro de 2020

This work was developed in collaboration with Instituto Português de Oncologia de Coimbra Francisco Gentil (IPOCFG), E.P.E.



This work was funded by Fundação para a Ciência e Tecnologia, through the PhD grant SFRH/BD/118929/2016.






**União Europeia**
Fundo Social Europeu

# Agradecimentos

Chega ao fim uma fase de enorme crescimento pessoal, que se traduz em momentos de felicidade, angústia, lágrimas e gargalhas, partilhados com aqueles que me rodeiam e pelos quais tenho um profundo sentimento de gratidão.

Começo por expressar o meu mais sincero agradecimento à Doutora Maria do Carmo Lopes, minha Orientadora e a grande responsável por este projeto, pela confiança, apoio e incentivo inexcedíveis. Obrigada por me ajudar a sonhar, acreditar, lutar e conquistar. Jamais esquecerei tudo o que fez e continua a fazer por mim. É uma verdadeira fonte de inspiração. Agradeço também ao Professor Paulo Crespo, pela disponibilidade e estímulo durante o doutoramento.

À Josefina, ao Miguel e ao Tiago, o meu profundo agradecimento por toda a ajuda, paciência e amizade. Vocês foram a minha segunda família ao longo destes anos e verdadeiramente essenciais nesta caminhada. Espero ter-vos deixado orgulhosos.

Obrigada à Ana, Ana Catarina, Carla, Carlos, Cármen, Catarina, Fábio, Fernanda, Liliana, Mary, João Pedro e Tozé, por me fazerem sentir parte desta equipa que é o Serviço de Física Médica do IPO de Coimbra.

Tenho também que agradecer ao conselho de administração do Instituto Português de Oncologia de Coimbra Francisco Gentil, E.P.E., por me ter acolhido e proporcionado as condições necessárias à realização do doutoramento. Agradeço igualmente à Agência Internacional de Energia Atómica e à Divisão de Física Médica da Sociedade Portuguesa de Física pelo apoio no projeto de auditoria nacional.

À Joanna Izewska, Eduard Gershkevitsh, Paulina Wesolowska, Pavel Kazantsev e a todos os físicos médicos dos centros de radioterapia que visitei, fica o meu reconhecimento pela disponibilidade, empenho e ajuda. Agradeço também à Doutora Joana Bastos e ao Professor Francisco Caramelo, pelo apoio prestado na análise estatística.

O meu muito obrigado à Brígida, D. Aida, D. Cila e Regina, pela atenção, conselhos e carinho. E à Eugénia, Helena, Inês e Liliana pela preocupação e incentivo. Obrigada por preservarem a nossa amizade, apesar das minhas ausências, por vezes, demasiado longas.

Ao Gonçalo, o meu profundo agradecimento pelo amor, compreensão, dedicação e apoio incondicional. Obrigada por tudo, que é tanto. Sem ti, tudo teria sido mais difícil.

Por fim, agradeço aos meus pais e irmão, a quem dedico esta tese, por me permitirem voar e serem o meu porto seguro. Ao Luís pequenino, à Laura, à vó Gelina, ao vô Tonho e à vó Ângela, deixo um obrigado por encherem a minha vida de sorrisos.

Muito obrigada!

"Never give up on something that you can't go a day without thinking about."
- Winston Churchill

# Abstract

Over the last decade, Intensity Modulated Radiation Therapy (IMRT), including Volumetric Arc Modulated Therapy (VMAT), has become a widespread treatment technique, given the associated clinical benefits. However, the inherent complexity leads to an increased potential for harmful errors to occur. For its safe and optimal use, a comprehensive quality assurance program must thus be implemented and a regular participation in external audits is recommended.

The International Atomic Energy Agency (IAEA), which has a vast experience in providing dosimetry services, has recently developed a national audit program to review the physics aspects of IMRT treatments. The established methodology includes the dosimetric verification of a head and neck IMRT plan through radiochromic film and ionization chamber measurements. It also comprises a set of tests designed for conventional linear accelerators to evaluate basic beam data as well as the machine performance. This audit can be adopted at the country level with the IAEA support.

In this context, the main objectives of this thesis were: 1) to carry out the IMRT audit in Portugal; 2) to contribute to extend the audit methodology, including its adaptation to Helical Tomotherapy (HT), adoption of plan quality and complexity metrics for plan assessment, and providing a practical guide for radiochromic film dosimetry.

The IMRT audit was conducted in Portugal in 2018 in collaboration with the IAEA and had 100% participation of the radiotherapy centres performing IMRT treatments. In the audit pilot centre, Instituto Português de Oncologia de Coimbra Francisco Gentil (IPOCFG), IMRT treatments are performed in a tomotherapy unit, which motivated the extension of the methodology to HT. The proposed adaptations included the definition of procedures to check the machine beam output, the treatment planning system (TPS) beam model and some small field dosimetry indicators.

Analysis of the quality and complexity of the audit plans, created by each participating institution using the local technologies/techniques, allows for an independent comparison of the followed approaches and sharing of experiences. The plans dosimetric quality was evaluated by using an in-house developed software tool, called SPIDERplan. To evaluate and compare the complexity of the audit IMRT/VMAT plans, multiple metrics were considered and a novel global plan complexity score was proposed. Additionally, the assessment of plan complexity was extended to HT, for which no indices had been proposed so far in the literature, through the definition and adaptation of multiple metrics.

Film dosimetry may be challenging, requiring the knowledge and implementation of complex procedures. A literature review on the practical aspects to take into account when setting up a film dosimetry system in clinical practice was done. It can serve as a guide for new users and audit teams that intend to conduct the IMRT audit at a national level. A comprehensive film dosimetry protocol for high dose verification of stereotactic treatments was also established, which can be considered in future extensions of the audit methodology to hypofractionated regimens.

**Keywords:** IMRT national audit, helical tomotherapy, plan quality, plan complexity, film dosimetry

# Resumo

Ao longo última década, a radioterapia de intensidade modulada (do inglês *intensity modulated radiation therapy* – IMRT), incluindo a arcoterapia volumétrica modulada (do inglês *volumetric modulated arc therapy* – VMAT), tornou-se uma técnica de tratamento amplamente usada. No entanto, a sua maior complexidade leva a que possam ocorrer mais erros relativamente às técnicas mais convencionais. As instituições devem por isso, implementar um programa de garantia de qualidade completo e participar regularmente em auditorias externas.

A Agência Internacional de Energia Atómica (IAEA), que tem um longo historial de experiência na disponibilização de auditorias dosimétricas, desenvolveu recentemente um programa de auditoria para tratamentos de IMRT. A metodologia adotada inclui a verificação de dose de um tratamento de IMRT de cabeça e pescoço usando filmes radiocrómicos e câmara de ionização. Contempla igualmente um conjunto de testes, especialmente definidos para aceleradores lineares convencionais, que visam a verificação da dosimetria básica e do desempenho da máquina de tratamento. Este programa pode ser adotado pelos diferentes Estados Membros, com o suporte da IAEA.

Neste contexto, esta tese teve como objetivos: 1) realizar a auditoria de IMRT em Portugal; 2) contribuir para a extensão da metodologia da auditoria, através da sua adaptação à Tomoterapia (HT), da inclusão de métricas de avaliação de qualidade e complexidade dos planos de tratamento e da proposta de um guião para a implementação de uma metodologia de dosimetria em filme.

A auditoria de IMRT foi conduzida em Portugal em 2018, em colaboração com a IAEA e contou com a participação de todos os centros de radioterapia que utilizavam, à data, IMRT na prática clínica. No centro piloto, Instituto Português de Oncologia de Coimbra Francisco Gentil (IPOCFG), estes tratamentos são executados na unidade de Tomoterapia, o que motivou a extensão da metodologia a este tipo de equipamento. A sua adaptação à HT incluiu a definição de procedimentos para verificar a calibração do feixe de tratamento, o modelo do feixe e indicadores de dosimetria básica de campos pequenos.

A análise da qualidade e complexidade dos planos de tratamento resultantes da auditoria permitiu uma avaliação independente das práticas locais e motivou a partilha de conhecimento. A qualidade dosimétrica dos planos foi avaliada usando uma ferramenta desenvolvida no IPOCFG – SPIDERplan. Para analisar e comparar a complexidade dos planos de IMRT/VMAT, foram consideradas várias métricas e proposto um *score* global. A avaliação da complexidade foi ainda estendida à HT, para a qual não existiam indicadores reportados na literatura, através da definição e adaptação de vários índices.

O uso de dosimetria em filme requer o conhecimento e execução de procedimentos muito detalhados. Assim, foi feita uma revisão bibliográfica dos aspetos a considerar na implementação de um sistema de dosimetria em filme. Esta revisão pode servir como guia a novos utilizadores ou equipas de auditores. Foi igualmente estabelecido um protocolo de dosimetria em filme para verificação de planos de radioterapia estereotáxica, que poderá ser considerado em extensões futuras da auditoria.

**Palavras-chave:** Auditoria nacional de IMRT, tomoterapia, avaliação da qualidade dos planos, avaliação da complexidade dos planos, dosimetria em filme

# Contents

# List of Figures

# List of Tables

# List of Abbreviations

| | |
|---|---|
| AAA | Analytical Anisotropic Algorithm |
| AAPM | American Association of Physicists in Medicine |
| ATP | Acceptance Testing Procedure |
| CAS | Cross-Axis Score |
| CBCT | Cone Beam Computed Tomography |
| CCC | Collapsed Cone Convolution |
| CCD | Charge Coupled Device |
| CHIP | Channel Independent Perturbation |
| CIRS | Computerized Imaging Reference Systems |
| CLS | Closed Leaf Score |
| CoP | Code of Practice |
| CP | Control Point |
| CT | Computed Tomography (CT) |
| DICOM | Digital Imaging and Communications in Medicine |
| DR | Dose Rate |
| DVH | Dose-Volume Histogram |
| EBT | External Beam Therapy |
| EM | Edge Metric |
| EPID | Electronic Portal Imaging Device |
| ESTRO | European SocieTy for Radiation and Oncology |
| FFF | Flattening Filter Free |
| FMC | Fluence Map Complexity |
| $f_{MSR}$ | Machine Specific Reference field |
| FWHM | Full Width at Half Maximum |
| FWQM | Full Width at Quarter Maximum |
| GS | Gold Standard |
| H&N | Head and Neck |
| HT | Helical Tomotherapy |
| HU | Hounsfield Unit |
| IAEA | International Atomic Energy Agency |
| IC | Ionization Chamber |
| ICRU | International Commission on Radiation Units and Measurements |
| IGRT | Image Guided Radiation Therapy |
| IMRT | Intensity Modulated Radiation Therapy |
| IPOCFG | Instituto Português de Oncologia de Coimbra Francisco Gentil |
| IROC | Imaging and Radiation Oncology Core |
| kV | Kilovoltage |
| kV-CBCT | Kilovoltage Cone Beam Computed Tomography |
| LED | Light Emission Diodes |
| Linac | Linear Accelerator |
| LiPCDA | Lithium-10,12-Pentacosdiynoate |
| LOT | Leaf Open Times |
| LOTV | Leaf Open Time Variability |
| LRA | Lateral Response Artifact |
| LT | Leaf Travel |
| L0NS | Leaves with Zero Neighbor Score |

| | |
|---|---|
| L1NS | Leaves with One Neighbor Score |
| MAD | Mean Asymmetry Distance |
| MC | Monte Carlo |
| MCS | Modulation Complexity Score |
| MF | Modulation Factor |
| MI | Modulation Index |
| MLC | Multileaf Collimator |
| MRI | Magnetic Resonance Imaging |
| MSR | Machine Reference Field |
| MU | Monitor Units |
| MV | Megavoltage |
| MV-CBCT | Megavoltage Cone Beam Computed Tomography |
| MVCT | Megavoltage Computed Tomography |
| NetOD | Net Optical Density |
| nPCS | Normalized Plan Complexity Score |
| NTCP | Normal Tissue Complication Probability |
| OAR | Organs-At-Risk |
| OD | Optical Density |
| OF | Output Factor |
| PC | Principal Component |
| PCA | Principal Component Analysis |
| PCS | Plan Complexity Score |
| PCSR | Plan-Class Specific Reference |
| PDD | Percentage Depth Dose |
| PDF | Probability Density Function |
| PET | Positron Emission Tomography |
| PI | Plan Irregularity |
| PM | Plan Modulation |
| PMMA | Polymethyl Methacrylate |
| PSTV | Plan Sinogram Time Variation |
| pT | Projection Time |
| PTV | Planning Target Volume |
| PTV_XXXX | Planning Target Volume where XXXX is the prescription dose in cGy |
| PV | Pixel Value |
| QA | Quality Assurance |
| QUATRO | Quality Assurance Team for Radiation Oncology |
| RED | Relative Electron Density |
| RGB | Red-Green-Blue |
| ROI | Region of Interest |
| RT | Radiotherapy |
| SAD | Source-Axis Distance |
| SAS | Small Aperture Score |
| SBRT | Stereotactic Body Radiotherapy |
| SD | Standard Deviation |
| SDD | Source-Detector Distance |
| SHANE | Shoulder Head and Neck End-to-end phantom |
| SRS | Stereotactic Radiosurgery |
| SS | Step & Shoot |

| | |
|---|---|
| SSD | Source-Skin Distance |
| SW | Sliding Window |
| WFF | With Flattening Filter |
| WHO | World Health Organization |
| TCP | Tumour Control Probability |
| TG | Task Group |
| TH | Threshold |
| TLD | Termoluminescent dosimeters |
| TPR | Tissue Phantom Ratio |
| TPS | Treatment Planning System |
| TRS | Technical Report Series |
| TT | Treatment Time |
| UV | Ultraviolet |
| VMAT | Volumetric Modulated Arc Therapy |
| 1D | One Dimensional |
| 2D | Two Dimensional |
| 3D | Three Dimensional |
| 3D-CRT | Three Dimensional Conformal Radiotherapy |

# CHAPTER 1

# Introduction

## 1.1 | Scope of the thesis

Cancer is the second leading cause of death worldwide, with 18.1 million new cases and 9.6 million deaths in 2018 [1]. In Portugal, 58199 diagnoses were registered in the same year, being the cause of approximately 25% of deaths [2,3]. Even if these numbers are already dramatic, cancer incidence is expected to keep rising [4]. Conventional cancer treatment modalities include radiotherapy, surgery and chemotherapy. It is estimated that about 50% of the incident cases require radiotherapy during the course of the treatment [5].

Radiotherapy technology has progressed considerably over the last decades and complex techniques such as intensity modulated radiation therapy (IMRT) are becoming commonly used, gradually replacing conventional techniques like three-dimensional conformal radiotherapy (3D-CRT). 3D-CRT uses radiation beams of uniform intensity that are shaped by a multileaf collimator (MLC) to the target volume geometrical configuration [6]. This is still the most used treatment technique in Portugal. In IMRT, the radiation beams intensity is modulated allowing for geometrical and dosimetric conformation, at the cost of increased complexity in the associated procedures. One of the most challenging tumour sites for IMRT is head and neck (H&N), with irregular target volume(s) and multiple organs-at-risk (OAR) in their close vicinity [7,8].

IMRT treatments can be delivered in several ways, either with a conventional linear accelerator (linac), including fixed gantry angle and rotational techniques (volumetric modulated radiation therapy, VMAT) or using other technologies like Helical Tomotherapy (HT) [8].

To ensure the safe and optimal implementation of IMRT and any other new treatment technique/machine in clinical practice, radiotherapy centres should participate in independent dosimetry audits [9]. However, national regulations in several countries, like Portugal [10], do not consider it as mandatory. Instead, participation in dosimetry audits is generally voluntary and driven by the professionals' motivation to perform high quality radiotherapy treatments [11]. The International Atomic Energy Agency (IAEA) has a long history in providing dosimetry services to its Member States, supporting also national activities through the development of new methodologies that can be adopted at the country level. In this line, being aware of the complexity associated with IMRT treatments, and the increased potential for errors, the IAEA has recently developed a national audit program for H&N IMRT. It simulates with an anthropomorphic phantom all steps of a patient treatment, from image acquisition and treatment planning to dose delivery, in an 'end-to-end' approach [12].

## 1.2 | Objectives of the thesis

In Portugal there were, in 2018, 24 radiotherapy centres with 55 treatment machines, including 52 linear accelerators, one Tomotherapy, one Cyberknife and one Gamma Knife [13]. Twenty centres had already introduced IMRT in clinical practice, and most of them (18/20) had never been audited for IMRT practices. Moreover, 16/20 centres did not participate in any audit initiative since the last national project on 3D-CRT in 2011/2012 [14]. In line with international recommendations, the first objective of the present project was to conduct the new IAEA national IMRT audit in Portugal and analyse the results. The audit methodology includes the dosimetric verification of an H&N IMRT plan created by each participating institution, through radiochromic film and ionization chamber measurements. A set of tests to evaluate TPS modelling of small fields, the MLC performance and machine beam output is also performed.

Given that the audit established tests are only valid for conventional linear accelerators and that the evaluation of the treatment plans quality is based on the analysis of a few dosimetric parameters, the second objective of the thesis was to contribute to extend the audit methodology, including:

i. Its adaptation to Helical Tomotherapy;

In HT, reference dosimetric conditions of conventional linacs do not apply, as the 10×10 $cm^2$ homogeneous field cannot be configured. Modified reference conditions have to be incorporated to verify the beam output calibration, using the concept of "machine specific reference field" defined in the IAEA and American Association of Physicists in Medicine (AAPM) code of practice TRS 483 [15]. In view of the HT specificities, small fields and the MLC performance also need to be handled in a different way.

ii. Adoption of plan quality and complexity metrics for plan assessment;

Considering the analysis of plan quality and complexity in the context of an IMRT dosimetry audit, where different treatment technologies/techniques are involved, may help to benchmark institutions with similar equipment, and eventually lead to an improvement of the planning strategies. According to the IMRT audit methodology, a plan is deemed satisfactory if it fulfils the dosimetric objectives/constraints, and no metrics are suggested to assess and compare the overall plans dosimetric quality. The total number of plan monitor units (MU) is indicated as a quick complexity measure, however, it does not give information on the parameters that contribute to an increased complexity.

iii. Providing a practical guide for radiochromic film dosimetry;

Radiochromic film allows to perform 2D measurements with a high spatial resolution, which is an important feature in IMRT plans verification. Yet, film dosimetry can be challenging, requiring the knowledge and implementation of complex procedures. The audit methodology does not provide detailed instructions on how to establish a reliable protocol and auditing teams that are not familiar with it, may find it difficult to implement.

## 1.3 | Organization of the thesis

This dissertation is organized into nine chapters:

Chapter 1 – Introduction: the scope and objectives of the thesis are defined.

Chapter 2 – Background Knowledge: basic and structuring concepts that allow to understand the developed work are reviewed, including an overview of IMRT, plan quality evaluation methods, plan complexity metrics, patient-specific quality assurance approaches, and dosimetry audits in radiotherapy.

Chapter 3 – IMRT national audit in Portugal: the IAEA audit program for on-site verification of H&N IMRT dose delivery has been implemented in Portugal in 2018, having the Instituto Português de Oncologia de Coimbra Francisco Gentil (IPOCFG) as the pilot centre. The national experience and results are presented in this chapter. A section was added to the end of the published paper to give details on the adaptations proposed for HT.

Chapter 4 – Evaluation of the complexity of treatment plans from a national IMRT/VMAT audit: the complexity of the IMRT and VMAT audit plans is quantified and compared through the calculation of multiple complexity metrics. A novel global plan complexity score is defined to summarize all the evaluated features through a single indicator. The dosimetric plan quality is assessed by using a software tool called SPIDERplan, developed at IPOCFG.

Chapter 5 – On the complexity of helical tomotherapy treatment plans: in this chapter, the complexity analysis is extended to tomotherapy. Multiple metrics are proposed to characterize the complexity of HT treatment plans created for different sites, corresponding to locally treated cases.

The global plan complexity score is used to compare the relative complexity of the entire set of plans.

Chapter 6 – A review on radiochromic film dosimetry for dose verification in high energy photon beams: a literature review on the practical aspects to take into account when implementing a film dosimetry system in clinical practice has been done. This can work as a guide for new users and audit teams that intend to conduct the IMRT audit at a national level.

Chapter 7 – A protocol for absolute dose verification of SBRT/SRS treatment plans using Gafchromic™ EBT-XD films: a film dosimetry protocol for high dose verification of stereotactic body radiotherapy (SBRT) and stereotactic radiosurgery (SRS) treatments has been established. It can possibly be considered in future extensions of the audit methodology.

Chapter 8 – Conclusions: the general conclusions of the developed work are presented.

Appendices – Complementary information to that provided at the end of Chapter 3 on the adaptations of the audit methodology to HT is provided:

- Appendix I – Application of the TRS 483 code of practice for reference and relative dosimetry in Tomotherapy: the recent IAEA/AAPM TRS 483 code of practice on dosimetry of small static fields and non-reference conditions was applied to helical tomotherapy for reference and relative dosimetry at IPOCFG.
- Appendix II – Independent verification of the pre-installed beam model in helical tomotherapy: HT is a pre-commissioned machine in the sense that it comes with a pre-installed beam model. Hence, it only needs to be verified by the customer as part of the commissioning process. The procedures adopted at IPOCFG for independent verification are described in this appendix.

These chapters correspond to six self-consistent papers (five published and one submitted to a journal) and a poster with the following references:

1. T Santos, MC Lopes, E Gershkevitsh, F Vinagre, D Faria, L Carita, et al. IMRT national audit in Portugal. Physica Medica 2019;65:128–36. Impact factor: 2.485 (Chapter 3)
2. T Santos, T Ventura, MC Lopes. Evaluation of the complexity of treatment plans from a national IMRT/VMAT audit – Towards a plan complexity score. Physica Medica 2020;70:75–84. Impact factor: 2.485 (Chapter 4)
3. T Santos, T Ventura, J Mateus, M Capela, MC Lopes. On the complexity of helical tomotherapy treatment plans. Journal of Applied Clinical Medical Physics 2020;21(7):107–18. Impact factor: 1.679 (Chapter 5)
4. T Santos, T Ventura, MC Lopes. A review on radiochromic film dosimetry for dose verification in high energy photon beams. Radiation Physics and Chemistry 2021;179:109217. Impact factor: 2.226 (Chapter 6)
5. T Santos, T Ventura, M Capela, J Mateus, MC Lopes. A protocol for absolute dose verification of SBRT/SRS treatment plans using Gafchromic™ EBT films. Submitted to Physica Medica in November 2020. (Chapter 7)
6. MC Lopes, T Santos, T Ventura, M Capela. Application of the TRS 483 code of practice for reference and relative dosimetry in tomotherapy. Medical Physics 2019;46(12):5799-5806. Impact factor: 3.317 (Appendix I)
7. MC Lopes, T Santos, T Ventura, M Capela. Independent verification of the pre-installed beam model in helical tomotherapy. Accepted for poster presentation at the International Conference on Advances in Radiation Oncology (ICARO-3), 16-19 February 2021 (virtual event). (Appendix II)

## 1.4 | References


[1] WHO. Factsheet WORLD. https://gco.iarc.fr/today/data/factsheets/cancers/39-All-cancers-fact-sheet.pdf (accessed October 12, 2020).

[2] PORTADA. Óbitos por algumas causas de morte. https://www.pordata.pt/Portugal/Óbitos+por+algumas+causas+de+morte+(percentagem)-758-235710 (accessed October 12, 2020).

[3] WHO. Factsheet PORTUGAL. https://gco.iarc.fr/today/data/factsheets/populations/620-portugal-fact-sheets.pdf (accessed October 12, 2020).

[4] WHO. Cancer Tomorrow. https://gco.iarc.fr/tomorrow/home (accessed October 12, 2020).

[5] Borras JM, Lievens Y, Dunscombe P, Coffey M, Malicki J, Corral J, et al. The optimal utilization proportion of external beam radiotherapy in European countries: An ESTRO-HERO analysis. Radiother Oncol 2015;116:38–44. doi:10.1016/j.radonc.2015.04.018.

[6] Hodapp N. The ICRU Report 83: prescribing, recording and reporting photon-beam intensity-modulated radiation therapy (IMRT). Strahlentherapie und Onkol 2012;188:97–9. doi:10.1007/s00066-011-0015-x.

[7] Teoh M, Clark CH, Wood K, Whitaker S, Nisbet A. Volumetric modulated arc therapy: a review of current literature and clinical use in practice. Br J Radiol 2011;84:967–96. doi:10.1259/bjr/22373346.

[8] Elith C, Dempsey SE, Findlay N, Warren-Forward HM. An introduction to the intensity-modulated radiation therapy (IMRT) techniques, tomotherapy, and VMAT. J Med Imaging Radiat Sci 2011;42:37–43. doi:10.1016/j.jmir.2010.11.005.

[9] Clark CH, Jornet N, Muren LP. The role of dosimetry audit in achieving high quality radiotherapy. Phys Imaging Radiat Oncol 2018;5:85–7. doi: 10.1016/j.phro.2018.03.009.

[10] Presidência do Conselho de Ministros. Decreto-Lei no108/2018 de 3 dezembro. Diário Da República, 1a Série 2018;232:5490–543.

[11] Izewska J, Lechner W, Wesolowska P. Global Availability of dosimetry audits in radiotherapy: The IAEA dosimetry audit networks database. Phys Imaging Radiat Onco. 2018;5:1-4. doi: 10.1016/j.phro.2017.12.002.

[12] Kazantsev P, Lechner W, Gershkevitsh E, Clark CH, Venencia D, Van Dyk J, et al. IAEA methodology for on-site end-to-end IMRT/VMAT audits: an international pilot study. Acta Oncol 2020;59:141–8. doi:10.1080/0284186X.2019.1685128.

[13] International Atomic Energy Agency (IAEA). DIRAC. https://dirac.iaea.org/Query/Countries (accessed October 7, 2020).

[14] Lopes MC, Cavaco A, Jacob K, Madureira L, Germano S, Faustino S, et al. Treatment planning systems dosimetry auditing project in Portugal. Phys Medica 2014;30:96–103. doi:10.1016/j.ejmp.2013.03.008.

[15] International Atomic Energy Agency (IAEA). Dosimetry of Small Static Fields Used in External Beam Radiotherapy, Technical Reports Series No. 483. Vienna: International Atomic Energy Agency; 2017.

# CHAPTER 2

## Background Knowledge

## 2.1 | Radiotherapy

Radiotherapy is a cancer treatment modality that uses ionizing radiation (photons, electrons, protons, heavy ions or neutrons) to administrate a therapeutic dose to tumour cells, whilst minimizing as much as possible healthy tissues damage. Radiotherapy modalities generally fall within two categories, depending on the radiation source location: brachytherapy and external beam radiotherapy. In brachytherapy, radioactive sources are placed inside or very closely to the tumour. In external beam radiotherapy, radiation beams commonly produced by a linear accelerator are delivered from outside the patient [1]. In this work, the term radiotherapy will be used to refer to external beam radiotherapy using megavoltage (MV) photon beams.

In external beam radiotherapy with megavoltage photon beams, three-dimensional conformal radiotherapy (3D-CRT) and intensity modulated radiation therapy (IMRT) are the most generally used delivery techniques. 3D-CRT involves the administration of multiple beams of uniform intensity that are shaped by a multileaf collimator (MLC) to match the target volume geometrical configuration. IMRT delivers beams of non-uniform intensity allowing for both geometric and dosimetric conformation [2]. Conventional radiotherapy treatments are performed on a daily basis, five days a week, during five to seven weeks, with a dose per fraction of 1.8 to 2 Gy. Stereotactic radiation therapy techniques where high doses are delivered in a single fraction or in only a few fractions are becoming standard of care for certain lesions, typically small and well-defined. Stereotactic techniques demand an accurate and reproducible localization of the target volume in space and time, which may be achieved through the use of rigid frames and image guided techniques. Stereotactic radiosurgery (SRS) refers to the treatment of intracranial lesions and stereotactic body radiation therapy (SBRT) to the treatment of extracranial tumours [3].

A radiotherapy treatment encompasses multiple steps, from clinical evaluation to post-treatment follow-up. The process starts, after a therapeutic decision has been taken, with a first appointment where the radiation oncologist prescribes the radiotherapy treatment scheme. Then, the patient immobilization to be adopted during treatment is defined and a computerized tomography (CT) scan of the patient is acquired, on which the radiation oncologist delineates the planning target volumes (PTV) to be irradiated and the organs-at-risk (OAR). To help in this task, the CT images may be co-registered with other imaging modalities such as magnetic resonance imaging (MRI) and positron emission tomography (PET). Based on the outlined structures and on the dose prescription to the PTV and tolerance dose criteria to the OARs, a treatment plan is created on a treatment planning system (TPS). After the treatment plan has been evaluated and approved by the radiation oncologist, there is a pre-treatment quality assurance (QA) verification to guarantee that it can be delivered as intended. Before each treatment fraction, the patient positioning is usually checked using an image guided radiation therapy (IGRT) modality. It typically involves the acquisition of two dimensional (2D) or three dimensional (3D) image sets that are compared with the reference planning images. During and after treatment, the patient is clinically followed by the radiation oncologist to evaluate his/her status and the treatment outcome in terms of tumour control and normal tissues complications.

## 2.2 | IMRT overview

In IMRT, the radiation is delivered using multiple beams of non-uniform intensity (fluence), which is achieved by subdividing each beam into beamlets. The use of a multileaf collimator enables the delivery of such complex patterns of modulated intensity. This allows for highly conformal dose distributions, with steep dose gradients surrounding the target volume, which

contribute to a better local control with fewer normal tissue complications. IMRT also permits to administrate a higher dose to regions of the target classified as at high risk disease and a lower dose to lower risk areas – simultaneous integrated boost. When used in combination with an IGRT modality for accurate localization, it enables the escalation of the target dose – hypofractionation. Yet, IMRT plans typically require an increased number of monitor units (MU) to deliver a stated dose, when compared with 3D-CRT. A higher number of MU is associated with a larger treatment time and volume of healthy tissues receiving low doses, which may lead to an increased risk of radiation induced malignancies. The complexity of most associated procedures, including basic dosimetry, planning process and treatment delivery is also increased. Consequently, QA processes required for IMRT are more demanding and time consuming [2,4,5].

IMRT treatments can be performed using a conventional linear accelerator (linac) or other technologies like Helical Tomotherapy. A conventional linac is an isocentric C-arm machine that allows to administrate radiation beams in multiple directions by the movement of the gantry, collimator and couch. The corresponding rotation axes intercept each other in a point called isocenter. The modulation of the beam is enabled by a multileaf collimator. MLCs are composed of many individual leaves arranged in pairs with dimensions from a few millimetres up to 1 cm at the isocenter level. The MLC leaves can move independently, allowing to create almost any shape [6]. Modern linacs are also equipped with one (or two) flat-panel detectors in the beam exit (or orthogonally), which by using the treatment beam (MV) or a kV x-ray source, enable for 2D/3D patient positioning verification prior to treatment. For 2D verification, at least two orthogonal images are acquired that are then compared to digitally reconstructed radiographies. For 3D verification, multiple 2D projections are obtained to reconstruct a volumetric dataset – CBCT (cone beam computed tomography) – that is then compared with the planning CT. If the treatment beam is used, the IGRT modality is known as MV-CBCT. Otherwise, if there is a kV x-ray source and a 2D detector system mounted perpendicularly to the treatment beam, the modality is called kV-CBCT [7]. There are other possible IGRT modalities like CT-on rails, where a CT scanner is installed in the treatment room [8].

On conventional linacs, IMRT can be delivered using fixed gantry angles – Step & Shoot and Sliding Window IMRT – or having the gantry rotating around the patient while the beam is on – Volumetric Modulated Arc Therapy (VMAT). In Step & Shoot IMRT, each incidence includes multiple static segments corresponding to different MLC aperture shapes, with the radiation being switched off between segments. In Sliding Window IMRT, the MLC leaf pairs continuously move while the beam is on [5], for a given incidence. VMAT allows treating partial or full 360° beam angle amplitude, with simultaneous variation of the gantry rotation speed, dose rate and MLC aperture shape. Like this, it is possible to treat the entire target volume in a single arc rotation, although in some complex cases more than one arc may be necessary [4]. The main advantages of VMAT in comparison to fixed gantry IMRT are indeed a significant reduction in treatment time and total number of MU [9–11].

Helical Tomotherapy consists in a 6 MV flattening filter free (FFF) linear accelerator mounted on a CT-like gantry, O-arm. In HT, the radiation beam produced by a compact linac is collimated to a “fan-beam”, with a width of 1, 2.5 or 5 cm, and a maximum lateral extension of 40 cm at isocenter, located at 85 cm from the radiation source. The beam intensity is modulated by a binary MLC composed by 64 tungsten leaves with a 0.625 cm width at isocenter, which can only be fully open or closed and whose movement is driven by a compressed air system. Delivery is done while the beam rotates around the patient and the couch translates through the gantry ring – helical IMRT. Both the gantry rotation and the couch translation speeds are constant throughout the treatment. In the gantry ring there is also an arc shaped detector mounted opposite to the linac that records the exit radiation signal. The recorded signal can be used for patient positioning verification through

the acquisition of megavoltage computed tomography images (MVCT), plan deliverability evaluation or machine quality assurance (QA). A great advantage of HT is that the length of the treatment region in the longitudinal direction is only limited by the maximum couch travel, up to 135 cm, so it can be used to treat extended target volumes or multiple lesions without field junctions [12,13].

IMRT treatment plans are designed using an inverse planning approach. The planner starts by defining a few parameters like the number of beam incidences (in static gantry IMRT), the number of arcs (in VMAT), or the field width (in HT) and then specifies dose-volume objectives and constraints for each PTV and OAR, as well as their relative importance. These descriptors are combined in a mathematical objective function that guides the TPS optimization algorithm, which iteratively determines the beamlet intensities most likely to give the desired dose distribution. During the optimization phase a manual fine tuning of the set parameters is usually required until an acceptable plan is obtained [2]. This is a trial and error process that can be quite time consuming and that is becoming smoothly replaced by automated solutions [14,15].

## 2.3 | Plan quality evaluation

A treatment plan is considered dosimetrically satisfactory when the dose prescription requirements set by the radiation oncologist are met. Yet, planners should aim for a solution that spares as much as possible the delineated OARs and healthy tissues as long as it uniformly irradiates the PTV with the prescription dose. In this process, a number of tentative plans are typically generated.

The most commonly used methods in clinical routine for dosimetric plan quality assessment are the visual inspection of the 3D dose distribution overlapped on CT planning images and the analysis of the dose-volume histograms (DVH) provided by the TPS for each delineated structure. The final evaluation and selection of the "optimal plan" from among competing solutions can thus be a cumbersome task due to the large and diverse amount of data to be considered. As a result, it is widely recognized that decision-making process is rather subjective, depending on the planner's and eventually radiation oncologist's knowledge and experience [16,17].

For making the plan selection process less ambiguous, some potential decision-support tools, more or less sophisticated, have been developed over the past decades. Dose-volume quality indices were first proposed in the 1990s by the Radiation Therapy Oncology Group to evaluate radiosurgery treatment plans, namely the coverage of the target volume by the prescription isodose, the conformity of the prescription isodose to the target volume and the homogeneity of the dose distribution [18]. Since then, various quality indices have been presented and more complete definitions have been described [19–30], considering, for instance, the organs-at-risk/normal tissue irradiation, the size and shape of the target volumes and the dose gradient outside the target volume. Still, such metrics have never been extensively used in the clinic for IMRT plan quality assessment, either due to limitations in their definition or complexity. An alternative approach includes the estimation of the outcome of the treatment by calculating some radiobiological indicators (using mathematical models) like tumour control probability (TCP) and normal tissue complication probability (NTCP). However, uncertainties related to the determination of the parameters needed in the biological modelling process still limit their applicability [31]. More recently, some software solutions have been presented. Ventura et al. [32], in 2016, have developed a tool called SPIDERplan that allows to assess and compare the quality of multiple plans through the combination of a graphical representation and the definition of a scoring index. The score functions, which are freely configurable, are generally based on the target coverage and OAR sparing, being

completely independent of the algorithm, treatment planning system or technique [17,32]. PlanIQ™ (Sun Nuclear Corporation, Melbourne, FL, USA) is a commercially available software that permits not only assessing plan quality based on customizable criteria, but also improving it [33]. PlanIQ™ provides, before planning, patient-specific feedback on the feasibility of the clinical objectives for each delineated structure, indicating how they can eventually be improved [34]. This *a priori* knowledge may contribute to reduce inter-planner variability [35].

The evaluation of plan quality in 'end-to-end' dosimetry audits is a good opportunity to assess what is achievable using different treatment technologies/techniques and lead to an improvement of the local planning methodologies through knowledge sharing.

## 2.4 | Plan complexity evaluation

IMRT allows to obtain highly conformal dose distributions and steep dose gradients via beam intensity modulation, which comes at the cost of intrinsic plan complexity [36]. High complex plans are associated with larger uncertainties in dose calculation and delivery, which may eventually compromise the accuracy and quality of the treatment [37].

Multiple metrics have been proposed and adapted to quantify the complexity degree of IMRT and VMAT plans [38–50]. The total number of MU per Gy (MU/Gy) is a simple and at the same time the most widely used indicator. A large MU/Gy is typically related to a higher complexity, as it may point to the presence of many small and irregularly shaped apertures [41,43,44]. Fluence map-based metrics like the Fluence Map Complexity (FMC) [47] and the Modulation Index (MI) [39], were the first proposed and evaluate the variations of intensity in fluence maps. The FMC presented by Llacer et al. in 2001, is defined as the normalized root square sum of the variation in intensity between each element and its lateral neighbours. The MI, proposed by Webb in 2003, accounts only for those differences between neighbours that are greater than a fraction of the standard deviation of the entire fluence map. Larger FMC and MI indicate a greater fluence heterogeneity that is associated with a higher level of complexity. However, these indices do not give indication on the parameters that contribute to obtain such complex patterns and the fluence maps may not be available [45,37].

Several metrics have been developed to directly evaluate treatment plan features, including the size, irregularity and location of MLC apertures, and modulation of machine parameters like the distance travelled by the MLC leaves as well as gantry speed and dose rate variations. The Modulation Complexity Score (MCS) presented by McNiven et al. in 2010 for Step & Shoot IMRT [42] and later modified by Masi et al. [48] for VMAT plans, takes simultaneously into account the shape, area and weight (MU) of each segment in the beam/plan. Other examples include, the Cross-Axis Score (CAS) [44] that assesses the amount of off-axis apertures, and the Plan Irregularity (PI) [43], which gives an indication on the apertures irregularity or narrowness. Variations of dose rate and gantry speed are usually analysed in VMAT plans, since they may be potentially demanding in terms of machine performance and MLC motion [49,50]. In this regard, Masi et al. [48] proposed an index, called Leaf travel (LT) to evaluate the mean MLC leaves motion per VMAT arc. Park et al. in 2013 [40], modified the MI presented by Webb et al. [39], in order to consider the variation of the MLC leaves position, gantry speed and dose rate in VMAT, by attributing a variable weight to each. Concerning other radiation therapy technologies, such as Tomotherapy, as far as we are aware of, no complexity metrics had been defined before those proposed in this thesis.

Complexity metrics have been mostly utilized to predict the likelihood of a plan to fail pre-treatment QA verification, with the final purpose of reducing the time allocated to QA measurement or eventually skip it. However, the majority of the studies were unable to identify thresholds that

would permit the separation between deliverable and non-deliverable beams/plans. These threshold values, even when found, strongly depend on the institution's technology for treatment planning and delivery, verification tools, evaluation metrics and acceptability criteria, so other centres cannot simply use them in clinical practice. Still, complexity metrics play an important role in treatment plans characterization and comparison, and may also contribute to establishing a trade-off between plan quality and complexity. As their definition is typically independent of the treatment technology/technique, such metrics can be useful in the context of multi-institutional dosimetry audits, allowing for comparison of treatment techniques, treatment machines, treatment planning systems and planner skills [51–53]. This may help to benchmark institutions with similar equipment, drive the medical physicists' attention to parameters that contribute to over-modulated plans, and eventually contribute to an improvement of the planning approaches.

## 2.5 | Patient-specific QA

There are multiple identified sources of error that may significantly compromise treatment deliverability, i.e. the agreement between the planned and delivered dose distributions, which are primarily related to TPS dose calculation, plan transfer, treatment machine performance, patient setup errors and internal motion during treatment. Some general aspects that may affect TPS dose calculations include modelling of small fields output factors, off-axis profiles, MLC leaf ends, tongue-and-groove effects, leaf/collimator transmission, and consideration of heterogeneity corrections. Regarding the treatment machine performance, MLC leaf position errors, MLC leaf speed, stability of the gantry rotation and couch motion, beam stability and the equipment age are some important factors to consider [54]. Given the potential impact of such errors in the patient treatment outcome, to introduce IMRT in clinical practice, a complete set of acceptance and commissioning tests should be performed and a comprehensive QA program designed and implemented [55,56]. The QA program must include periodic tests of the treatment machine performance and image guidance tools as well as patient-specific verifications to evaluate the deliverability of each plan either prior to treatment – pre-treatment QA verification – or during treatment – in-vivo dosimetry [57–59].

The purpose of pre-treatment QA verification is usually two-folded. First (if applicable), it aims at verifying if each treatment plan is properly transferred from the TPS to the treatment machine, and second, if it can be delivered as intended. There are multiple verification tools available that include: ionization chamber and other point dose detectors, radiochromic film, 2D and 3D detector arrays, treatment machine image detectors, and independent monitor unit (or time) verification softwares [54].

### 2.5.1 | Point dose verification - ionization chamber

Ionization chambers are used as a reference dosimeter for point dose verification due to their long term stability, linear response to absorbed dose and traceability to a primary calibration standard. The treatment plan is typically recalculated in a phantom and a cylindrical ionization chamber is placed in a homogeneous dose region. As ionization chambers present some volume averaging, measurement in high dose gradient regions can lead to large deviations between planned and determined doses [54].

### 2.5.2 | 2D dose verification

**Radiochromic film**

A dosimetry system composed by radiochromic films and a flatbed document scanner enables 2D dose measurements with a submilimeter spatial resolution. Radiochromic films are self-developing and have many attractive characteristics, such as: near-water equivalent composition, weak energy dependence in radiotherapy photon MV range, angular independence, and consistent response in an extended range of doses. As for ionization chamber measurements, radiochromic films are usually placed at a plane of interest inside a phantom [60].

**Planar detector arrays**

2D detector arrays such as MapCHECK2 (Sun Nuclear Corporation, Melbourne, FL, USA) and MatriXX (IBA Dosimetry, Schwarzenbrook, Germany) consist of small ionization chambers or diodes arranged in a matrix with different spacing between them. These devices were firstly used for per-beam verification of fixed-gantry IMRT, with the beam incidences perpendicular to the measurement plane. When utilized for true composite verifications (either multiple IMRT incidences or rotational techniques), their planar design does not allow to fully measure lateral beams and there is a directional dependence of the response (anisotropy), which is more pronounced for diode-based arrays [61,62]. The planar detectors can also be attached to the gantry to measure the beam fluence hence eliminating angular dependence.

### 2.5.3 | 3D dose verification

**"3D" detector arrays**

There are some detector array systems that allow for 3D dose verification, like OCTAVIUS 4D (PTW, Freiburg, Germany), Delta$^4$ phantom (Scandidos, Uppsala, Sweden) and ArcCHECK (Sun Nuclear Corporation, Melbourne, FL, USA). OCTAVIUS 4D (PTW, Freiburg, Germany) consists in a polystyrene cylindrical phantom, which rotates synchronously with the gantry, coupled with a planar ionization chamber array. The planar array can have a spatial resolution from 2.5 mm (OCTAVIUS Detector 1000SRS) to 1 cm (OCTAVIUS Detector 729) [63]. Delta$^4$ phantom (Scandidos, Uppsala, Sweden) is constituted by 1069 diodes disposed along two orthogonal planes in a cylindrical PMMA phantom. The diodes spacing is 5 mm in the central area (6×6 cm$^2$) and 1 cm in the outer area. The measurement is synchronized with the irradiation [64]. ArcCHECK (Sun Nuclear Corporation, Melbourne, FL, USA) records the signal of 1386 diodes, spaced by 1 cm, and embedded as a helical grid on a PMMA cylindrical phantom [63]. A problem pointed to these devices is the limited spatial resolution which may influence the pre-treatment QA results due to undersampling effects [65].

### 2.5.4 | Image detectors: EPID and Tomotherapy detector

The image detectors already existing in the treatment units enable to measure the exit fluence during irradiation before and during treatment. The signal recorded by electronic portal imaging devices (EPID) in linacs and the image detector in Tomotherapy can be used to reconstruct the delivered dose distribution on a phantom/patient CT set, using manufacturer's software or third party solutions [66].

### 2.5.5 | Independent monitor unit (or time) calculation

Systems to perform an independent calculation of the treatment monitor units (time) to deliver a certain dose to a given point(s) have been used for many years [67]. More sophisticated alternatives include the calculation of the 3D dose distribution in the patient anatomy using an independent calculation algorithm [68].

### 2.5.6 | Comparison of measured and calculated dose distributions

The Gamma index, introduced by Low et al. in 1998 [69], is one of the most used metrics to assess the agreement between measured and planned dose distributions. This method allows to compare the dose distributions in a quantitative way, combining percent dose difference and distance-to-agreement in one parameter – the gamma index [69]. Basically, the two dose distributions to be compared are assigned as the reference and the evaluated distributions. For each point in the reference dose distribution, $r_R$, the distance $\Delta r$ and the dose difference $\Delta D$ to each point in the evaluated distribution $r_E$ are determined, and the gamma index calculated from:

$$\Gamma(r_R, r_E) = \sqrt{\frac{\Delta r^2(r_R, r_E)}{\delta r^2} + \frac{\Delta D^2(r_R, r_E)}{\delta D^2}} \tag{2.1}$$

where $\delta r$ is the distance-to-agreement and $\delta D$ the dose difference criteria [69,70]. The gamma index corresponds to:

$$\gamma = min\{\Gamma(r_R, r_E)\}, \forall\{r_E\} \tag{2.2}$$

If $\gamma \leq 1$, the acceptance criteria are met. When the dose difference tolerance is relative to the local/point dose, the analysis is known as local gamma analysis. When it is relative to a global dose value in a low gradient region such as the maximum predicted TPS dose in a dose matrix plane, it is named global gamma analysis. It is common practice to report the evaluation results through the percentage of points for which the gamma values meet the acceptance criteria ($\gamma \leq 1$), the so-called gamma passing rate.

An AAPM report on methodologies and tolerance limits for IMRT pre-treatment QA verification has been recently published [54]. Among other recommendations, it is suggested that the agreement between planned and measured dose distributions should be assessed using global gamma analysis with 3%/2 mm 10% dose threshold criteria, and an acceptance level of 95%. The threshold is used to exclude dose regions with no/limited clinical significance that can bias the analysis results [54]. Still, several studies reported that the gamma passing rates can be misleading as revealed to be insensitive to clinically significant dosimetric errors, being DVH-based metrics suggested as an alternative or at least a complement to the gamma analysis [71–73]. In this line, analysis software modules [74–76] have been developed to reconstruct the 3D dose distribution on the patient CT images and calculate the DVH for each delineated structure from the performed measurements.

The conclusions regarding treatment plans deliverability may be different, depending on the verification systems, analysis methods, criteria and tolerances used [54]. Regular participation in dosimetry audits is thus highly advised for independent review of the local practices.

## 2.6 | Dosimetry audits

Independent dosimetry audits play an important role in patient treatment quality and safety. Audits have the potential to identify issues and resolve them, reducing the probability of harmful errors to occur. Audits can also support the safe implementation of new techniques and technologies, and promote knowledge sharing at a national and/or international level by benchmarking centres with similar equipment [77]. Indeed, the IAEA stresses the importance of every radiotherapy centre equipped with new machines and those that are going to introduce new treatment techniques in clinical practice, participate in dosimetry audits before starting treating patients, and regularly after that [78]. Moreover, a recent European Directive (2013/59 Euratom) [79] recommends that new radiological procedures should be audited. Independent dose audits are also mandatory in many multi-institutional clinical trials in radiotherapy to ensure that participants deliver accurate doses and so the reported results are not biased [80–82].

There are different types of dosimetry audits, performed either at a national level or large scale, and including both postal and on-site visits. The dosimetry audit programs vary from machine beam calibration measurements in reference conditions to 'end-to-end' verifications. 'End-to-end' approaches simulate with an anthropomorphic phantom all steps of a patient treatment including, image acquisition, treatment planning, and dose delivery [77].

The IAEA has a vast experience in promoting dosimetry audits in radiotherapy, involving both postal services and on-site visits. A postal audit program has been conducted in collaboration with the World Health Organization (WHO) to check the calibration of high energy photon beams for over 50 years [83], using passive luminescent dosimeters. The dosimeters are mailed to the participants that irradiate them to a specified dose under reference conditions and then returned to the IAEA, which analyses and compares the intended to the measured dose. There are other international organizations that conduct this type of audits, as, for example, Imaging and Radiation Oncology Core group in Houston (IROC-H) [84] in the United States, and the Equal-Estro laboratory (EQUAL) in Europe [1] [85].

A national 'end-to-end' audit program of 3D-CRT treatment techniques was developed by the IAEA in 2008 [86]. This audit was carried out in Portugal in 2011/2012 [87], with the IAEA assistance. Some methodologies have already been established for IMRT verification at a national/regional level [88–93] or on a large scale [85,94–96]. An independent audit of the local pre-treatment QA methods has been conducted in the Netherlands [90], where the same set of pre-defined IMRT/VMAT clinical plans was considered for dose calculation in all institutions. In about 20% of the evaluated measurements, the results of the audit and the local QA were in disagreement. The IROC – Houston QA Center has implemented a remote program for clinical trials credentialing of H&N IMRT treatments, using a semi-anthropomorphic phantom [94]. Significant improvements in the results have been achieved over the years, but still about 10% of the irradiations fail to meet the tolerances of ±7% dose difference for point dose measurement with thermoluminescent dosimeters (TLD) and 85% global gamma analysis (7% dose difference/4 mm distance-to-agreement) for film [97]. The IAEA has also developed a national audit program to review the physics aspects of H&N IMRT treatments, following an 'end-to-end' approach. Portugal was one of the first countries implementing this audit at a national level. The audit methodology comprises the dosimetric verification of an H&N IMRT plan created by each participating institution, and a set of tests to check small field dosimetry, MLC performance, and the machine beam output [98].

[1] It is now an independent laboratory not linked to the European SocieTy for Radiation and Oncology (ESTRO)

A few dosimetry audits of SBRT and SRS treatments, which are complex and high risk techniques, have already been set-up [99–101]. However, further efforts and developments are needed to widespread such methodologies.

There are also more comprehensive audit programs, like the Quality Assurance Team for Radiation Oncology (QUATRO). This program, provided by the IAEA, reviews and evaluates the overall quality of all radiotherapy related aspects at an institution, including infrastructure, organization, documentation, equipment, procedures, radiation protection aspects, staffing levels and training. The audit is carried out by a multidisciplinary team, including a radiation oncologist, a medical physicists and a radiotherapy technologist [102].

## 2.7 | References


[1] Berman; AT, Kharofa Jordan. Introduction to Radiation Oncology. https://www.astro.org/uploadedfiles/affiliates/arro/future_residents/introtoro.pdf (accessed November 3, 2020).

[2] Hodapp N. The ICRU Report 83: prescribing, recording and reporting photon-beam intensity-modulated radiation therapy (IMRT). Strahlentherapie und Onkol 2012;188:97–9. doi:10.1007/s00066-011-0015-x.

[3] Wilke L, Andratschke N, Blanck O, Brunner TB, Combs SE, Grosu A-L, et al. ICRU report 91 on prescribing, recording, and reporting of stereotactic treatments with small photon beams: Statement from the DEGRO/DGMP working group stereotactic radiotherapy and radiosurgery. Strahlenther Onkol 2019;195:193–8. doi:10.1007/s00066-018-1416-x.

[4] Teoh M, Clark CH, Wood K, Whitaker S, Nisbet A. Volumetric modulated arc therapy: a review of current literature and clinical use in practice. Br J Radiol 2011;84:967–96. doi:10.1259/bjr/22373346.

[5] Elith C, Dempsey SE, Findlay N, Warren-Forward HM. An introduction to the intensity-modulated radiation therapy (IMRT) techniques, tomotherapy, and VMAT. J Med Imaging Radiat Sci 2011;42:37–43. doi:10.1016/j.jmir.2010.11.005.

[6] Podgorsak E. Radiation Oncology Physics: A Handbook for Teachers and Students. Med Phys 2006;33:1920. doi:10.1118/1.2201870.

[7] Gupta T, Narayan CA. Image-guided radiation therapy: Physician's perspectives. J Med Phys 2012;37:174–82. doi:10.4103/0971-6203.103602.

[8] Fontenot JD, Alkhatib H, Garrett JA, Jensen AR, McCullough SP, Olch AJ, et al. AAPM Medical Physics Practice Guideline 2.a: Commissioning and quality assurance of X-ray-based image-guided radiotherapy systems. J Appl Clin Med Phys 2014;15:3–13. doi:10.1120/jacmp.v15i1.4528.

[9] Yin L, Wu H, Gong J, Geng J-H, Jiang F, Shi A-H, et al. Volumetric-modulated arc therapy vs. c-IMRT in esophageal cancer: a treatment planning comparison. World J Gastroenterol 2012;18:5266–75. doi:10.3748/wjg.v18.i37.5266.

[10] Eppinga W, Lagerwaard F, Verbakel W, Slotman B, Senan S. Volumetric modulated arc therapy for advanced pancreatic cancer. Strahlenther Onkol 2010;186:382–7. doi:10.1007/s00066-010-2094-5.

[11] Weber DC, Peguret N, Dipasquale G, Cozzi L. Involved-node and involved-field volumetric modulated arc vs. fixed beam intensity-modulated radiotherapy for female patients with early-stage supra-diaphragmatic Hodgkin lymphoma: a comparative planning study. Int J Radiat Oncol Biol Phys 2009;75:1578–86. doi:10.1016/j.ijrobp.2009.05.012.

[12] Fenwick JD, Tomé WA, Soisson ET, Mehta MP, Rock Mackie T. Tomotherapy and other innovative IMRT delivery systems. Semin Radiat Oncol 2006;16:199–208. doi:10.1016/j.semradonc.2006.04.002.

[13] Kupelian P, Langen K. Helical tomotherapy: image-guided and adaptive radiotherapy. Front Radiat Ther Oncol 2011;43:165–80. doi:10.1159/000322420.

[14] Breedveld S, Storchi PRM, Voet PWJ, Heijmen BJM. iCycle: Integrated, multicriterial beam angle, and profile optimization for generation of coplanar and noncoplanar IMRT plans. Med Phys 2012;39:951–63. doi:10.1118/1.3676689.

[15] Krayenbuehl J, Norton I, Studer G, Guckenberger M. Evaluation of an automated knowledge based treatment planning system for head and neck. Radiat Oncol 2015;10:226. doi:10.1186/s13014-015-0533-2.

[16] Nelms BE, Robinson G, Markham J, Velasco K, Boyd S, Narayan S, et al. Variation in external beam treatment plan quality: An inter-institutional study of planners and planning systems. Pract Radiat Oncol 2012;2:296–305. doi:10.1016/j.prro.2011.11.012.

[17] Ventura T, Dias J, Khouri L, Netto E, Soares A, Ferreira BC, et al. Clinical validation of a graphical method for radiation therapy plan quality assessment. Radiat Oncol 2020;15:64. doi:10.1186/s13014-020-01507-5.

[18] Shaw E, Kline R, Gillin M, Souhami L, Hirschfeld A, Dinapoli R, et al. Radiation Therapy Oncology Group: radiosurgery quality assurance guidelines. Int J Radiat Oncol Biol Phys 1993;27:1231–9. doi:10.1016/0360-3016(93)90548-a.

[19] Cheung FWK, Law MYY. A novel conformity index for intensity modulated radiation therapy plan evaluation. Med Phys 2012;39:5740–56. doi:10.1118/1.4742848.

[20] Leung LHT, Kan MWK, Cheng ACK, Wong WKH, Yau CC. A new dose-volume-based Plan Quality Index for IMRT plan comparison. Radiother Oncol 2007;85:407–17. doi:10.1016/j.radonc.2007.10.018.

[21] Miften MM, Das SK, Su M, Marks LB. A dose-volume-based tool for evaluating and ranking IMRT treatment plans. J Appl Clin Med Phys 2004;5:1–14.

[22] Menhel J, Levin D, Alezra D, Symon Z, Pfeffer R. Assessing the quality of conformal treatment planning: a new tool for quantitative comparison. Phys Med Biol 2006;51:5363–75. doi:10.1088/0031-9155/51/20/019.

[23] Lomax NJ, Scheib SG. Quantifying the degree of conformity in radiosurgery treatment planning. Int J Radiat Oncol Biol Phys 2003;55:1409–19. doi:10.1016/s0360-3016(02)04599-6.

[24] Baltas D, Kolotas C, Geramani K, Mould RF, Ioannidis G, Kekchidi M, et al. A conformal index (COIN) to evaluate implant quality and dose specification in brachytherapy. Int J Radiat Oncol Biol Phys 1998;40:515–24. doi:10.1016/s0360-3016(97)00732-3.

[25] Paddick I. A simple scoring ratio to index the conformity of radiosurgical treatment plans. Technical note. J Neurosurg 2000;93 Suppl 3:219–22. doi:10.3171/jns.2000.93.supplement.

[26] Jin X, Yi J, Zhou Y, Yan H, Han C, Xie C. A new plan quality index for nasopharyngeal cancer SIB IMRT. Phys Medica 2014;30:122–7. doi:10.1016/j.ejmp.2013.03.007.

[27] van't Riet A, Mak AC, Moerland MA, Elders LH, van der Zee W. A conformation number to quantify the degree of conformality in brachytherapy and external beam irradiation: application to the prostate. Int J Radiat Oncol Biol Phys 1997;37:731–6. doi:10.1016/s0360-3016(96)00601-3.

[28] Akpati H, Kim C, Kim B, Park T, Meek A. Unified dosimetry index (UDI): a figure of merit for ranking treatment plans. J Appl Clin Med Phys 2008;9:2803. doi:10.1120/jacmp.v9i3.2803.

[29] Wagner TH, Bova FJ, Friedman WA, Buatti JM, Bouchet LG, Meeks SL. A simple and reliable index for scoring rival stereotactic radiosurgery plans. Int J Radiat Oncol Biol Phys 2003;57:1141–9. doi:10.1016/S0360-3016(03)01563-3.

[30] Yan L, Xu Y, Chen X, Xie X, Liang B, Dai J. A new homogeneity index definition for evaluation of radiotherapy plans. J Appl Clin Med Phys 2019;20:50–6. doi:10.1002/acm2.12739.

[31] Marks LB, Yorke ED, Jackson A, Ten Haken RK, Constine LS, Eisbruch A, et al. Use of normal tissue complication probability models in the clinic. Int J Radiat Oncol Biol Phys 2010;76:S10–9. doi:10.1016/j.ijrobp.2009.07.1754.

[32] Ventura T, Lopes MC, Ferreira BC, Khouri L. SPIDERplan: A tool to support decision-making in radiation therapy treatment plan assessment. Reports Pract Oncol Radiother 2016;21:508–16. doi:10.1016/j.rpor.2016.07.002.

[33] Corporation SN. PlanIQ 2020. https://www.sunnuclear.com/uploads/documents/datasheets/PlanIQ_D032017.pdf (accessed September 21, 2020).

[34] Ahmed S, Nelms B, Gintz D, Caudell J, Zhang G, Moros E, et al. A method for a priori estimation of best feasible DVH for organs-at-risk: Validation for head and neck VMAT planning. Med Phys 2017;44:5486–97. doi:10.1002/mp.12500.

[35] Alves N, Ventura T, Mateus J, Capela M, Lopes MC. Do a priori expectations of plan quality offset planning variability in head and neck IMRT? Radiat Phys Chem 2020;168:108580. doi:10.1016/j.radphyschem.2019.108580.

[36] Crowe SB, Kairn T, Middlebrook N, Sutherland B, Hill B, Kenny J, et al. Examination of the properties of IMRT and VMAT beams and evaluation against pre-treatment quality assurance results. Phys Med Biol 2015;60:2587–601. doi:10.1088/0031-9155/60/6/2587.

[37] Hernández V, Hansen C, Widesott L, Bäck A, Canters R, Fusella M, et al. What is plan quality in radiotherapy? The importance of evaluating dose metrics, complexity, and robustness of treatment plans. Radiother Oncol 2020. doi:10.1016/j.radonc.2020.09.038.

[38] Younge KC, Matuszak MM, Moran JM, McShan DL. Penalization of aperture complexity in inversely planned volumetric modulated arc therapy. Med Phys 2012;39:7160–70. doi:10.1118/1.4762566.

[39] Webb S. Use of a quantitative index of beam modulation to characterize dose conformality: Illustration by a comparison of full beamlet IMRT, few-segment IMRT (fsIMRT) and conformal unmodulated radiotherapy. Phys Med Biol 2003;48:2051–62. doi:10.1088/0031-9155/48/14/301.

[40] Park JM, Park SY, Kim H, Ho Kim J, Carlson J, Ye SJ. Modulation indices for volumetric modulated arc therapy. Phys Med Biol 2014;59:7315–40. doi:10.1088/0031-9155/59/23/7315.

[41] Mohan R, Arnfield M, Tong S, Wu Q, Siebers J. The impact of fluctuations in intensity patterns on the number of monitor units and the quality and accuracy of intensity modulated radiotherapy. Med Phys 2000;27:1226–37. doi:10.1118/1.599000.

[42] McNiven AL, Sharpe MB, Purdie TG. A new metric for assessing IMRT modulation complexity and plan deliverability. Med Phys 2010;37:505–15. doi:10.1118/1.3276775.

[43] Du W, Cho SH, Zhang X, Hoffman KE, Kudchadker RJ. Quantification of beam complexity in intensity-modulated radiation therapy treatment plans. Med Phys 2014;41. doi:10.1118/1.4861821.

[44] Crowe SB, Kairn T, Kenny J, Knight RT, Hill B, Langton CM, et al. Treatment plan complexity metrics for predicting IMRT pre-treatment quality assurance results. Australas Phys Eng Sci Med 2014;37:475–82. doi:10.1007/s13246-014-0274-9.

[45] Chiavassa S, Bessieres I, Edouard M, Mathot M, Moignier A. Complexity metrics for IMRT and VMAT plans: a review of current literature and applications. Br J Radiol 2019;92:20190270. doi:10.1259/bjr.20190270.

[46] Antoine M, Ralite F, Soustiel C, Marsac T, Sargos P, Cugny A, et al. Use of metrics to quantify IMRT and VMAT treatment plan complexity: A systematic review and perspectives. Phys Medica Eur J Med Phys 2019;64:98–108. doi:10.1016/j.ejmp.2019.05.024.

[47] Llacer J, Solberg TD, Promberger C. Comparative behaviour of the Dynamically Penalized Likelihood algorithm in inverse radiation therapy planning. Phys Med Biol 2001;46:2637. doi:10.1088/0031-9155/46/10/309.

[48] Masi L, Doro R, Favuzza V, Cipressi S, Livi L. Impact of plan parameters on the dosimetric accuracy of volumetric modulated arc therapy. Med Phys 2013;40. doi:10.1118/1.4810969.

[49] Nicolini G, Clivio A, Cozzi L, Fogliata A, Vanetti E. On the impact of dose rate variation upon RapidArc (R) implementation of volumetric modulated arc therapy. Med Phys 2011;38:264–71. doi:10.1118/1.3528214.

[50] Miura H, Tanooka M, Fujiwara M, Takada Y, Doi H, Odawara S, et al. Predicting Delivery Error Using a DICOM-RT Plan for Volumetric Modulated Arc Therapy. Int J Med Physics, Clin Eng Radiat Oncol 2014;3:82–7. doi: 10.4236/ijmpcero.2014.32013.

[51] Hernandez V, Saez J, Pasler M, Jurado-Bruggeman D, Jornet N. Comparison of complexity metrics for multi-institutional evaluations of treatment plans in radiotherapy. Phys Imaging Radiat Oncol 2018;5:37–43. doi: 10.1016/j.phro.2018.02.002.

[52] Glenn MC, Hernandez V, Saez J, Followill DS, Howell RM, Pollard-Larkin JM, et al. Treatment plan complexity does not predict IROC Houston anthropomorphic head and neck phantom performance. Phys Med Biol 2018;63. doi: 10.1088/1361-6560/aae29e.

[53] McGarry CK, Agnew CE, Hussein M, Tsang Y, McWilliam A, Hounsell AR, et al. The role of complexity metrics in a multi-institutional dosimetry audit of VMAT. Br J Radiol 2016;89. doi: 10.1259/bjr.20150445.

[54] Miften M, Olch A, Mihailidis D, Moran J, Pawlicki T, Molineu A, et al. Tolerance limits and methodologies for IMRT measurement-based verification QA: Recommendations of AAPM Task Group No. 218. Med Phys 2018;45:e53–83. doi: 10.1002/mp.12810.

[55] Smilowitz JB, Das IJ, Feygelman V, Fraass BA, Kry SF, Marshall IR, et al. AAPM Medical Physics Practice Guideline 5.a.: Commissioning and QA of Treatment Planning Dose Calculations - Megavoltage Photon and Electron Beams. J Appl Clin Med Phys 2015;16:14–34. doi:10.1120/jacmp.v16i5.5768.

[56] Ezzell GA, Burmeister JW, Dogan N, LoSasso TJ, Mechalakos JG, Mihailidis D, et al. IMRT commissioning: multiple institution planning and dosimetry comparisons, a report from AAPM Task Group 119. Med Phys 2009;36:5359–73. doi:10.1118/1.3238104.

[57] Mijnheer B. 206 Guidelines for the verification of IMRT. Report of an ESTRO working group. Radiother Oncol 2005;76. doi:10.1016/S0167-8140(05)81183-1.

[58] Mans A, Schuring D, Arends MP, Vugts CAJM, Wolthaus JWH, Lotz HT, et al. The NCS code of practice for the quality assurance and control for volumetric modulated arc therapy. Phys Med Biol 2016;61:7221–35. doi:10.1088/0031-9155/61/19/7221.

[59] Wal E, Wiersma J, Ausma AH, Cuijpers J, Tomsej M, Bos L, et al. Code of Practice for the Quality Assurance and Control for Intensity Modulated Radiotherapy Report 22 of the Netherlands Commission on Radiation Dosimetry. 2013.

[60] Niroomand-Rad A, Chiu-Tsao S-T, Grams MP, Lewis DF, Soares CG, Van Battum LJ, et al. Radiochromic Film Dosimetry: An Update to TG-55. doi:10.1002/mp.14497.

[61] Han Z, Ng S, Bhagwat M, Lyatskaya Y, Zygmanski P. Evaluation of MatriXX for IMRT and VMAT dose verifications in peripheral dose regions. Med Phys 2010;37:3704–14. doi:10.1118/1.3455707.

[62] Jursinic PA, Sharma R, Reuter J. MapCHECK used for rotational IMRT measurements: step-and-shoot, TomoTherapy, RapidArc. Med Phys 2010;37:2837–46. doi:10.1118/1.3431994.

[63] Urso P, Lorusso R, Marzoli L, Corletto D, Imperiale P, Pepe A, et al. Practical application of Octavius-4D: Characteristics and criticalities for IMRT and VMAT verification. J Appl Clin Med Phys 2018;19:517–24. doi:10.1002/acm2.12412.

[64] Bedford JL, Lee YK, Wai P, South CP, Warrington AP. Evaluation of the Delta4 phantom for IMRT and VMAT verification. Phys Med Biol 2009;54:N167-76. doi:10.1088/0031-9155/54/9/N04.

[65] Hussein M, Rowshanfarzad P, Ebert MA, Nisbet A, Clark CH. A comparison of the gamma index analysis in various commercial IMRT/VMAT QA systems. Radiother Oncol 2013;109:370–6. doi:10.1016/j.radonc.2013.08.048.

[66] Chung E, Kwon D, Park T, Kang H, Chung Y. Clinical implementation of Dosimetry CheckTM for TomoTherapy(®) delivery quality assurance. J Appl Clin Med Phys 2018;19:193–9. doi:10.1002/acm2.12480.

[67] Sellakumar P, Arun C, Sanjay SS, Ramesh SB. Comparison of monitor units calculated by radiotherapy treatment planning system and an independent monitor unit verification software. Phys Med 2011;27:21–9. doi:10.1016/j.ejmp.2010.01.006.

[68] Kodama T, Saito Y, Hatanaka S, Hariu M, Shimbo M, Takahashi T. Commissioning of the Mobius3D independent dose verification system for TomoTherapy. J Appl Clin Med Phys 2019;20:12–20. doi:10.1002/acm2.12572.

[69] Low DA, Harms WB, Mutic S, Purdy JA. A technique for the quantitative evaluation of dose distributions. Med Phys 1998;25:656–61. doi:10.1118/1.598248.

[70] Hussein M, Clark C, Nisbet A. Challenges in calculation of the gamma index in radiotherapy – Towards good practice. Phys Medica 2017;36:1–11. doi:10.1016/j.ejmp.2017.03.001.

[71] Nelms BE, Zhen H, Tomé W a. Per-beam, planar IMRT QA passing rates do not predict clinically relevant patient dose errors. Med Phys 2011;38:1037–44. doi:10.1118/1.3544657.

[72] Kruse JJ. On the insensitivity of single field planar dosimetry to IMRT inaccuracies. Med Phys 2010;37:2516–24. doi:10.1118/1.3425781.

[73] Zhen H, Nelms BE, Tomé WA. Moving from gamma passing rates to patient DVH-based QA metrics in pretreatment dose QA. Med Phys 2011;38:5477. doi:10.1118/1.3633904.

[74] Hauri P, Verlaan S, Graydon S, Ahnen L, Klöck S, Lang S. Clinical evaluation of an anatomy-based patient specific quality assurance system. J Appl Clin Med Phys 2014;15:4647. doi:10.1120/jacmp.v15i2.4647.

[75] Song JH, Shin H-J, Kay CS, Son SH. Dosimetric Verification by Using the ArcCHECK System and 3DVH Software for Various Target Sizes. PLoS One 2015;10:e0119937. doi: 10.1371/journal.pone.0119937

[76] Pimthong J, Kakanaporn C, Tuntipumiamorn L, Laojunun P, Iampongpaiboon P. Commissioning and validation of COMPASS system for VMAT patient specific quality assurance. J Phys Conf Ser 2016;694:12025. doi:10.1088/1742-6596/694/1/012025.

[77] Clark CH, Jornet N, Muren LP. The role of dosimetry audit in achieving high quality radiotherapy. Phys Imaging Radiat Oncol 2018;5:85–7. doi:10.1016/j.phro.2018.03.009.

[78] International Atomic Energy Agency. Accuracy Requirements and Uncertainties in Radiotherapy. Vienna: IAEA Human Health Series; 2016. doi:10.1038/sj.bjc.6604224.

[79] European Parliament. Council Directive 2013/59/Euratom of 5 December 2013 laying down basic safety standards for protection against the dangers arising from exposure to ionising radiation, and repealing Directives 89/618/Euratom, 90/641/Euratom, 96/29/Euratom, 97/43/Euratom a. Off J Eur Commun L13 2014:1–73. doi:10.3000/19770677.L_2013.124.eng.

[80] Ibbott GS, Haworth A, Followill DS. Quality assurance for clinical trials. Front Oncol 2013;3:311. doi:10.3389/fonc.2013.00311.

[81] Weber DC, Poortmans PMP, Hurkmans CW, Aird E, Gulyban A, Fairchild A. Quality assurance for prospective EORTC radiation oncology trials: the challenges of advanced technology in a multicenter international setting. Radiother Oncol 2011;100:150–6. doi:10.1016/j.radonc.2011.05.073.

[82] Melidis C, Bosch W, Izewska J, Fidarova E, Zubizarreta E, Ishikura S, et al. Radiation therapy quality assurance in clinical trials-Global Harmonisation Group. Radiother Oncol 2014;111. doi:10.1016/j.radonc.2014.03.023.

[83] Izewska J, Bokulic T, Kazantsev P, Wesolowska P, Merwe D. 50 Years of the IAEA/WHO postal dose audit programme for radiotherapy: what can we learn from 13756 results? Acta Oncol (Madr) 2020;59:1–8. doi:10.1080/0284186X.2020.1723162.

[84] Alvarez P, Kry SF, Stingo F, Followill D. TLD and OSLD dosimetry systems for remote audits of radiotherapy external beam calibration. Radiat Meas 2017;106:412–5. doi: 10.1016/j.radmeas.2017.01.005.

[85] Veres, A.; Hallet JX. Equal-Estro experience in dosimetry audits in advanced techniques of radiotherapy – the tomotherapy example. SSDL Newsl No 70 2019:29–31.

[86] International Atomic Energy Agency. TECDOC 1583: Commissioning of Radiotherapy Treatment Planning Systems. IAEA. Vienna: 2008.

[87] Lopes MC, Cavaco A, Jacob K, Madureira L, Germano S, Faustino S, et al. Treatment planning systems dosimetry auditing project in Portugal. Phys Medica 2014;30:96–103. doi: 10.1016/j.ejmp.2013.03.008.

[88] Miri N, Lehmann J, Legge K, Zwan BJ, Vial P, Greer PB. Remote dosimetric auditing for intensity modulated radiotherapy: A pilot study. Phys Imaging Radiat Oncol 2017;4:26–31. doi: 10.1016/j.phro.2017.11.004.

[89] Lafond C, Chiavassa S, Bertaut C, Boussion N, Chapel N, Chapron L, et al. A multi-institutional dosimetry audit of rotational and static intensity-modulated radiotherapy. Phys Medica 2019;32:664–70. doi:10.1016/j.ejmp.2016.04.008.

[90] Seravalli E, Houweling AC, Battum L Van, Raaben TA, Kuik M, de Pooter JA, et al. Auditing local methods for quality assurance in radiotherapy using the same set of predefined treatment plans. Phys Imaging Radiat Oncol 2018;5:19–25. doi: 10.1016/j.phro.2018.01.002.

[91] Jurado-Bruggeman D, Hernández V, Sáez J, Navarro D, Pino F, Martínez T, et al. Multi-centre audit of VMAT planning and pre-treatment verification. Radiother Oncol 2017;124:302–10. doi: 10.1016/j.radonc.2017.05.019.

[92] Clark CH, Hansen VN, Chantler H, Edwards C, James H V, Webster G, et al. Dosimetry audit for a multi-centre IMRT head and neck trial. Radiother Oncol 2009;93:102–8. doi: 10.1016/j.radonc.2009.04.025.

[93] Clark CH, Hussein M, Tsang Y, Thomas R, Wilkinson D, Bass G, et al. A multi-institutional dosimetry audit of rotational intensity-modulated radiotherapy. Radiother Oncol 2014;113:272–8. doi:10.1016/j.radonc.2014.11.015.

[94] Molineu A, Hernandez N, Nguyen T, Ibbott G, Followill D. Credentialing results from IMRT irradiations of an anthropomorphic head and neck phantom. Med Phys 2013;40. doi: 10.1118/1.4773309.

[95] Lechner W, Wesolowska P, Azangwe G, Arib M, Gabriel V, Alves L, et al. A multinational audit of small field output factors calculated by treatment planning systems used in radiotherapy. Phys Imaging Radiat Oncol 2018;5:58–63. doi:10.1016/j.phro.2018.02.005.

[96] Izewska J, Wesolowska P, Azangwe G, Followill DS, Thwaites DI, Arib M, et al. Testing the methodology for dosimetry audit of heterogeneity corrections and small MLC-shaped fields: Results of IAEA multi-center studies. Acta Oncol 2016;55:909–16. doi:10.3109/0284186X.2016.1139180.

[97] Carson ME, Molineu A, Taylor PA, Followill DS, Stingo FC, Kry SF. Examining credentialing criteria and poor performance indicators for IROC Houston's anthropomorphic head and neck phantom. Med Phys 2016;43:6491. doi: 10.1118/1.4967344.

[98] Kazantsev P, Lechner W, Gershkevitsh E, Clark CH, Venencia D, Van Dyk J, et al. IAEA methodology for on-site end-to-end IMRT/VMAT audits: an international pilot study. Acta Oncol 2020;59:141–8. doi:10.1080/0284186X.2019.1685128.

[99] Distefano G, Lee J, Jafari S, Gouldstone C, Baker C, Mayles H, et al. A national dosimetry audit for stereotactic ablative radiotherapy in lung. Radiother Oncol 2017;122:406–10. doi:10.1016/j.radonc.2016.12.016.

[100] Dimitriadis A, Tsang Y, Thomas R, Palmer A, Eaton D, Lee J, et al. Multi-institutional dosimetric delivery assessment of intracranial stereotactic radiosurgery on different treatment platforms. Radiother Oncol 2020. doi:10.1016/j.radonc.2020.05.024.

[101] Lambrecht M, Melidis C, Sonke J-J, Adebahr S, Boellaard R, Verheij M, et al. Lungtech, a phase II EORTC trial of SBRT for centrally located lung tumours - a clinical physics perspective. Radiat Oncol 2016;11:7. doi:10.1186/s13014-015-0567-5.

[102] Izewska J, Coffey M, Scalliet P, Zubizarreta E, Santos T, Vouldis I, et al. Improving the quality of radiation oncology: 10 years' experience of QUATRO audits in the IAEA Europe Region. Radiother Oncol 2018;126:183–90. doi:10.1016/j.radonc.2017.09.011.

# CHAPTER 3

# IMRT national audit in Portugal

Physica Medica, 2019, volume 65, pages 128-136

T Santos[a,b], MC Lopes[b], E Gershkevitsh[c], F Vinagre[d], D Faria[e], L Carita[f], M Pontes[g], S Vieira[h], E Poli[i], S Faustino[j], F Ribeiro[k], M Trindade[l], F Ponte[m], C Marcelino[n], C Batista[o], S Oliveira[p], R Figueira[q], J Lencart[r], E Gallego-Diaz[s], K Jacob[t], S Brás[u], R Pirraco[v], J Izewska[w]

[a] Departamento de Física, Universidade de Coimbra, Coimbra, Portugal
[b] Serviço de Física Médica, IPOCFG, E.P.E., Coimbra, Portugal
[c] North Estonia Medical Centre, Tallinn, Estonia
[d] CHUC, E.P.E., Coimbra, Portugal
[e] Lenitudes Medical Center and Research, Sta Maria da Feira, Portugal
[f] IPOLFG, E.P.E., Lisboa, Portugal
[g] JCS - Centro Oncológico Dra Natália Chaves, Carnaxide, Portugal
[h] Fundação Champalimaud, Lisboa, Portugal
[i] CHULN-Hospital de Sta Maria, Lisboa, Portugal
[j] Hospital da Luz, Lisboa, Portugal
[k] CHBM, E.P.E., Barreiro, Portugal
[l] CHTMAD, E.P.E., Vila Real, Portugal
[m] Instituto CUF Porto, Porto, Portugal
[n] Lenicare, Évora, Portugal
[o] JCS - Clínica de Radioncologia de Santarém, Santarém, Portugal
[p] JCS - Clínica de Radioncologia do Algarve, Faro, Portugal
[q] CHUSJ, E.P.E., Porto, Portugal
[r] IPOPFG, E.P.E., Porto, Portugal
[s] JCS - Clínica de Radioncologia da Madeira, Madeira, Portugal
[t] Hospital CUF Descobertas, Lisboa, Portugal
[u] JCS - Clínica de Radioncologia Madalena Paiva, Açores, Portugal
[v] Hospital de Braga, Braga, Portugal
[w] Dosimetry and Medical Radiation Physics Section, International Atomic Energy Agency, Vienna, Austria

## Abstract

**Purpose:** The IAEA newly developed "end-to-end" audit methodology for on-site verification of IMRT dose delivery has been carried out in Portugal in 2018. The main goal was to evaluate the physical aspects of the head and neck (H&N) cancer IMRT treatments. This paper presents the national results.

**Methods:** All institutions performing IMRT treatments in Portugal, 20 out of 24, have voluntarily participated in this audit. Following the adopted methodology, a Shoulder, Head and Neck End-to-End phantom (SHANE) – that mimics an H&N region, underwent all steps of an IMRT treatment, according to the local practices. The measurements using an ionization chamber placed inside the SHANE phantom at four reference locations (three in PTVs and one in the spinal cord) and an EBT3 film positioned in a coronal plane were compared with calculated doses. FilmQA Pro software was used for film analysis.

**Results:** For ionization chamber measurements, the percent difference was within the specified tolerances of ±5% for PTVs and ±7% for the spinal cord in all participating institutions. Considering film analysis, gamma passing rates were on average 96.9%±2.9% for a criterion of 3%/3 mm, 20% threshold, all above the acceptance limit of 90%.

**Conclusions:** The national results of the H&N IMRT audit showed a compliance between the planned and the delivered doses within the specified tolerances, confirming no major reasons for concern. At the same time, the audit identified factors that contributed to increased uncertainties in the IMRT dose delivery in some institutions resulting in recommendations for quality improvement.

**Keywords:** intensity modulated radiation therapy, on-site dosimetry audit, head and neck treatment

## 3.1 | Introduction

Intensity modulated radiation therapy (IMRT), including volumetric modulated arc therapy (VMAT), has become a standard treatment technique given the associated benefits [1]. However, in comparison to 3-dimensional conformal radiotherapy (3D-CRT), IMRT is more demanding in both planning and delivery, requiring a strict quality assurance (QA) program. As a complement, many international organizations [2–8], including the International Atomic Energy Agency (IAEA), recommend that every institution that intends to introduce such complex techniques in clinical practice should participate in an independent dosimetry audit.

The IAEA has a vast experience in promoting dosimetry audits in radiotherapy to its Member States, involving both postal services and on-site visits. A postal audit program has been conducted in collaboration with the World Health Organization to check the calibration of high energy photon beams for 50 years now [9]. The methodology for a national on-site end-to-end audit of 3D-CRT treatment techniques using an anthropomorphic thorax phantom [10] was developed by the IAEA in 2008. This audit has been carried out in eight countries in Europe [11], including Portugal [12]. With the introduction of IMRT/VMAT in clinical practice, development of audit procedures to ensure the safe use of this technique became necessary. Thus, some methodologies have already been established and used at a national/regional level [7,13–17] or on a large scale [18–21]. An independent audit of the local patient-specific QA systems has been recently conducted in the Netherlands [14] where the same set of pre-defined IMRT/VMAT clinical plans was considered in all institutions for dose computation. In about 20% of the evaluated measurements, the results of audit and the local QA were in disagreement. The Radiological Physics Center in Houston (now called Imaging and Radiation Oncology Core, IROC – Houston QA Center) has implemented a remote program for clinical trials credentialing of Head and Neck (H&N) IMRT treatments, using a semi-anthropomorphic phantom [19]. Thermoluminescent dosimeters (TLD) and radiochromic film are used to assess the agreement between the calculated and measured doses. The current acceptability criteria is ±7% dose difference for TLD and a passing rate of 85% in global gamma analysis (7% dose difference/4 mm distance-to-agreement) for film. Significant improvements in the results have been achieved over the years, but still about 10% of the irradiations fail to meet the tolerances [22].

In large scale settings with several hundred facilities to be audited, remote audits are typically preferred due to their cost-effectiveness. As part of a multi-step remote audit system, the IAEA has disseminated a methodology for postal end-to-end IMRT/VMAT audits using TLDs and film in a specially developed solid phantom that is being implemented by some national dosimetry audit networks [23]. However, on-site visits have demonstrated advantages over remote postal audits – results are typically available without delay and any issues or discrepancies can be discussed between the auditor and the auditee, and possibly resolved during the visit; more detailed observations can be made, e.g. regarding phantom positioning and the performance of local QA as well as the overall process of IMRT delivery at the institution. Usually, postal audits are more restricted in terms of the audit scope and program, whereas on-site visits allow more flexibility, e.g. additional modalities or delivery techniques can be audited as needed and if time allows. As main drawbacks, on-site visits involve substantial travel costs and the auditor's time.

Recognizing the complexity associated with the IMRT procedures and being aware of the advantages of on-site visits in some national settings, the IAEA has recently developed a national audit program to review the physics aspects of H&N IMRT treatments, following an end-to-end approach [24]. A specially designed anthropomorphic phantom named Shoulder, Head and Neck End-to-End (SHANE – Computerized Imaging Reference Systems (CIRS) Inc., Norfolk, Virginia,

USA) [25] and a DICOM set of pre-defined contours are used to represent a patient with a nasopharynx tumour. The leading objective of the audit is to verify on-site that the complete radiotherapy treatment chain from computed tomography (CT) scanning to dose delivery leads to the desired results within the defined tolerances. The feasibility of the methodology was verified through a multicentre pilot study in a variety of clinical settings and the feedback by participants was carefully considered to fine-tune it [26]. This audit has been carried out in Portugal in 2018 with the IAEA assistance. The Medical Physics Division of the Portuguese Physics Society was designated the national auditing organization and Instituto Portugues de Oncologia Coimbra (IPOCFG, E.P.E.) the pilot centre. The national experience and results are presented in this paper.

## 3.2 | Materials and Methods

### 3.2.1 | National characterization

In Portugal, in 2018, there were 24 radiotherapy institutions equipped with 55 treatment machines including, 52 linear accelerators (linacs), one Tomotherapy, one Cyberknife and one Gamma Knife. By the time of the audit, twenty institutions had already introduced IMRT in clinical practice, with 14/20 having more than 2 years of experience in H&N IMRT. IMRT treatments represented about one third of the total external beam radiotherapy treatments in the country. H&N patients corresponded on average to 20% of the total number of patients treated with IMRT.

All 20 institutions voluntarily participated in the audit with one equipment set (linac/TPS) and IMRT technique. The treatment units comprised 19 linacs (14 Varian and 5 Elekta) and one Tomotherapy. From these, 14 treatment machines were less than 10 years old and 10 of them had been installed in the past 5 years. The oldest linac had more than 15 years. Regarding the nominal beam energy, 18/20 institutions used 6 MV and two 6FFF MV (Tomotherapy and one linac). The IMRT delivery techniques were: volumetric modulated arc therapy – VMAT (15), sliding window (3), step & shoot (1) and helical IMRT (1). The treatment planning systems included 13 Eclipse (Varian), 5 Monaco (Elekta), 1 XiO (Elekta) and 1 VoLO (Accuray). And finally, the dose calculation algorithms were: AAA (12), AcurosXB (1), Monte Carlo (5), fast superposition (1), and CCC superposition (1). Dose calculation grid resolution varied between 1 and 3 mm: 1 mm (1), 2 mm (5), 2.5 mm (11) and 3 mm (3).

### 3.2.2 | Audit phases

Before starting the audit program, a kick-off workshop was organized at the pilot centre to present the adopted methodology and discuss its implementation. The project encompassed different phases that are described in the following subsections.

**Pre-visit activities**

Some pre-visit activities were required to be performed by the participating institutions prior to the on-site visit. Aiming at speeding up the treatment planning phase, each institution had to create a preliminary IMRT plan based on a pre-visit CT dataset of the SHANE phantom and the corresponding pre-delineated structures, provided by the IAEA. The structures set included three planning target volumes (PTV_7000, PTVn1_6000 and PTVn2_5400) and four organs-at-risk (spinal cord, brainstem, left parotid and right parotid). For treatment planning, each institution considered the local treatment technique and their typical calculation specifications for H&N (dose calculation algorithm, grid resolution, inclusion of the treatment couch, etc.). To guide the optimization process, dose-volume constraints were provided (Table S3.1 – Supplementary data).

In addition, participants were asked to calculate on their TPS, output factors (OF) for 5 MLC-shaped fields (10×10 $cm^2$, 6×6 $cm^2$, 4×4 $cm^2$, 3×3 $cm^2$ and 2×2 $cm^2$) to compare with the IROC-Houston QA Centre's reference dataset [27,28]. Tolerances of ±3% for the 2×2 $cm^2$ field and ±2% for larger fields were considered [24]. Inplane and crossplane profiles for a MLC-shaped 2×2 $cm^2$ field, at source-axis distance (SAD) of 100 cm, depth of 10 cm, were also calculated in the TPS. Field sizes (defined as the normalized dose profile full width at half maximum – FWHM) and penumbra widths (20-80%) were recorded.

**Pilot centre audit**

The audit at the pilot centre was performed in the presence of an IAEA expert at the Tomotherapy unit. Before the SHANE irradiations, a dose comparison was done between the IAEA calibrated ionization chamber (IC) and a pilot centre's dosimetry system composed of a TM31010 Semiflex 0.125 cc ionization chamber and UNIDOS E electrometer (PTW-Freiburg, Germany) in the machine specific reference conditions for Tomotherapy [29]. This aimed at ensuring the metrological quality of the pilot centre's dosimetry system so that it could be used for audit measurements at participating centres. The result of the dose comparison was within 0.2%.

**On-site visits**

Following the IAEA audit methodology, the national auditor travelled through the institutions with the phantom, the audit dosimetry system, a barometer, a thermometer and a box with Gafchromic EBT3 films (Ashland Inc., Covington, Kentucky, USA) from a single batch, appropriately cut, numbered and marked. All on-site measurements were performed using the audit dedicated equipment. After each visit, the equipment integrity and long-term stability of the dosimetry system composed by the ionization chamber and electrometer were checked.

Each on-site visit took two days, the first for phantom scanning and treatment planning and the second for dose measurements.

a) CT scan of the SHANE phantom

The on-site visit started by performing a CT scanning of the SHANE phantom – on-site planning CT set – following the local CT scanning protocol for H&N patients. The phantom was positioned on the couch and aligned without using any immobilization devices such as head rests or thermoplastic masks as per the audit instructions. In the shoulders region of SHANE phantom seven reference materials are embedded – lung inhale, lung exhale, water vial, soft tissue, spinal cord, trabecular bone and cortical bone – with certified relative electron density (RED)/mass density values. After transferring the on-site planning CT set to the local TPS, a CT to RED/mass density curve was built to be compared with the one stored in the TPS [24]. Differences between curves of ±5 Hounsfield Unit (HU) for water and ±20 HU for all other materials were considered acceptable, as per the IAEA TRS-430 recommendations for TPS verification [30].

b) Treatment planning phase

The pre-visit CT set was co-registered with the on-site planning CT using rigid registration and the structures were copied over. The volume of each delineated structure was checked against a reference set of expected volume values, provided by the IAEA, to ensure that no major changes had occurred. The adopted tolerances were established based on the analysis of data resulting from a multicentre study to test the feasibility of the treatment planning exercise [31]. The pre-visit plan was then transferred (copied or saved as template) to the on-site planning CT set and re-optimized. The provided set of contours included four structures that represented the ionization chamber

volume surrogating the measurement reference points. The mean dose over those structures was registered.

Patient-specific QA of the IMRT created plan was performed by the local medical physics team using their equipment, analysis metrics, and acceptability criteria as for a typical H&N cancer patient. Some institutions considered more than one verification method. A summary of the used systems is presented in Table 3.1.

**Table 3.1** – Local verification systems used to validate the H&N IMRT audit plan.

| **System** | **Manufacturer** | **No. of institutions** |
|---|---|---|
| Image detector dosimetry | EPID dosimetry (Varian Medical Solutions, Palo Alto, California, USA)<br>Dosimetry Check (LifeLine Software Inc., Austin, TX, USA) | 8 |
| MatriXX | IBA Dosimetry, Schwarzenbruck, Germany | 4* |
| ArcCHECK | SunNucelar Corporation, Melbourne, FL, USA | 4 |
| Octavius II + PTW 729 array | PTW-Freiburg, Germany | 1 |
| Octavius 4D + PTW 1500 array + IC measurements | PTW-Freiburg, Germany | 3 |
| EBT3 film + IC measurement | Ashland Inc., Covington, Kentucky, USA | 2 |
| IC measurements | | 2** |
| Independent MU/dose calculation | RadCalc (LifeLine Software Inc., Austin, TX, USA)<br>Mobius3D (Varian Medical Solutions) | 3 |

*In 1 institution the MatriXX array was used in combination with a slab phantom and in 3 with MultiCube; in one of them it was also attached to the gantry head, and COMPASS software was used to reconstruct 3D dose; **Ionization chamber measurement in combination with slab solid water phantom.

c) On-site measurements

The on-site measurements were performed on the second day. Besides the dosimetric verification of the newly created H&N IMRT plan, measurements also included: i) a test to evaluate MLC performance; ii) irradiation of an EBT3 film with a 2×2 $cm^2$ MLC-shaped field as configured in the pre-treatment phase; iii) beam output verification; iv) irradiation of dose reference strips for film calibration, along with verification using the audit dosimetry system. The MLC test used either Electronic Portal Imaging Device (EPID) or film, considering an irradiation pattern as per IAEA specifications (5 strips of all MLC length, 3 cm gap between strips and the minimum achievable leaf opening). Film was placed isocentrically on a solid slab phantom (SAD=100 cm) with at least 2 cm of build-up. EPID was positioned as close as possible to the isocenter.

Daily output of the machine was checked against the reference dose calibration modelled in the TPS (10×10 $cm^2$ field, SAD=100 cm, depth=10 cm) with the audit dosimetry system. The measured dose was determined according to the TRS 398 [32] dosimetry protocol and compared with the calculated dose as follows:

$$Deviation\ [\%] = 100 \times (D_{cal} - D_{meas})/D_{meas} \quad (3.1)$$

where $D_{cal}$ is the TPS calculated and $D_{meas}$ is the measured dose. A tolerance of ±2% was considered.

Dosimetric verification of the created H&N IMRT treatment plan was done through ionization chamber measurements performed using the audit dosimetry system and EBT3 film irradiation. The ionization chamber was positioned in one of the four measurement reference locations, and at least

two irradiations were performed per reference point. The calculated and measured doses were compared using equation (3.1) but taking $D_{cal}$ as the corrected calculated dose for daily output variation – $D_{cal^*}$. The established tolerances were ±5% for PTVs and ±7% for the spinal cord [26]. An EBT3 film was then placed in a coronal plane of SHANE and given three treatment fractions as per IAEA methodology [24]. After measurements with SHANE, the reference strips were irradiated for film calibration.

### Post-visit analysis

All films irradiated during the on-site visit were analysed centrally by the national auditing team at the pilot centre. Calibration and measurement films were scanned at least 48 hours after irradiation using a flatbed scanner Epson Expression 10000 XL (Seiko Epson Corporation, Japan). To guarantee a stabilized scanner response, 16 empty scans were performed after a warm-up time of at least 30 minutes. A glass compression plate was put on films top to ensure film flatness. Red-Green-Blue (RGB) images were acquired in transmission mode, at 48 bits colour depth, 72 dpi, landscape orientation and with all colour corrections turned off.

For EBT3 film calibration, 10 strips were irradiated to known doses (0, 1, 2, 3, 4, 5, 6, 7, 8 and 9 Gy) in reference conditions (10×10 cm$^2$ field, SAD=100 cm, depth=10 cm) in a 6 MV photon mode at the pilot centre. A calibration curve for each channel (red, green and blue) was obtained from the average pixel values of a central region of interest of 1×1 cm$^2$.

For audit films processing, the conversion from pixel values to dose was performed using the generic calibration curve updated by two reference strips at zero dose and 110% over the maximum TPS calculated dose in the film plane (irradiated in each institution) [33]. The agreement between the film dose distribution and TPS calculated dose was evaluated with FilmQA Pro software (Ashland Inc., Covington, Kentucky, USA) using triple channel dosimetry, mimicking IAEA practice [23,26]. 2D global gamma analysis was performed, with normalization done to a high dose low gradient region. The tolerance limit was a gamma passing rate of 90%, for a criterion of 3% of global dose/3 mm distance-to-agreement with 20% dose threshold [26]. TPS dose distribution was considered the reference, meaning that the film dose distribution was rescaled to the resolution of the TPS dose grid. To minimize the impact of grid size on gamma analysis, the calculated dose distribution in the film plane was exported with a resolution of about 1 mm in all institutions, where possible.

To investigate the correlation between ionization chamber (IC) dose differences and gamma passing rates for film analysis, an average percent deviation was calculated considering the four measurement points, as follows:

$$\frac{1}{4}\left[\left(\frac{|D^*_{cal}-D_{meas}|}{D_{meas}}\right)_{PTV\_7000}+\left(\frac{|D^*_{cal}-D_{meas}|}{D_{meas}}\right)_{PTV_{n1}\_6000}+\left(\frac{|D^*_{cal}-D_{meas}|}{D_{meas}}\right)_{PTV_{n2}\_5400}+\left(\frac{|D^*_{cal}-D_{meas}|}{D_{meas}}\right)_{SpinalCord}\right] \quad (3.2)$$

However, when using equation (3.2) the spinal cord is given the same weight as PTVs, ignoring that the corresponding measurement point is located in a low dose high gradient region. Therefore, in a second analysis, the last term (SpinalCord) of equation (3.2) was excluded from the average IC percent deviation.

The EBT3 films irradiated with the 2×2 cm$^2$ field after being scanned and converted to dose maps, were analysed with RIT113 software v5.1 (Radiological Imaging Technology Inc., Colorado Springs, CO, USA), the clinically used software at the pilot centre. The measured field size (defined

as the normalized dose profile FWHM) and the penumbra width (20-80%) were compared with those calculated in the pre-visit phase. The tolerances were ±2 mm for the field size and ±3 mm for the penumbra width.

MLC tests performed on EPID were analysed using the freeware software Pylinac implemented as a web app (https://assuranceqa.herokuapp.com/). Films were analysed using the FilmQA Pro software. The maximum leaf bias and the median positioning error were checked. The tolerances were ±1 mm for the maximum leaf bias and ±0.5 mm for the median positioning error [34].

## 3.3 | Results

### 3.3.1 | Small field dosimetry and MLC performance tests

For the output factors calculated on TPS using 6 MV photon mode, the percent differences to IROC-Houston QA Centre's reference data [27,28] were determined. The average values are presented in Table 3.2 as a function of field size. On average, the audited TPSs overestimated OFs in comparison to the reference IROC-Houston dataset. Nevertheless, the differences were generally within the tolerances, with three institutions having a deviation higher than 2% for 3×3 $cm^2$ field size and four exceeding the tolerance of 3% for the 2×2 $cm^2$ field.

The measured field size and the penumbra widths obtained from analysis of the film irradiated with a 2×2 $cm^2$ MLC-shaped field were compared with the ones calculated on TPS. Differences were within ±2 mm for both the field size and penumbras in all institutions.

MLC performance test results were within ±0.5 mm for the individual leaf pairs positioning bias in all institutions. The maximum positioning error was on average 0.19±0.11 mm (maximum of 0.49 mm) and the average of the median bias was 0.05±0.03 mm (maximum of 0.10).

**Table 3.2 –** Summary of the relative difference between OFs calculated by TPS and reference dataset.

| | **Field size (cm×cm)** | | | |
|---|---|---|---|---|
| | **6×6** | **4×4** | **3×3** | **2×2** |
| **Overall (N=18)** | | | | |
| mean | 0.1% | 0.6% | 1.4% | 2.1% |
| standard deviation (SD) | 0.6% | 0.6% | 0.7% | 0.9% |
| N exceeding the tolerance | - | - | 3 | 4 |
| **Varian linac – Eclipse TPS (N=12)** | | | | |
| mean | 0.1% | 0.9% | 1.6% | 2.3% |
| standard deviation | 0.1% | 0.3% | 0.4% | 0.7% |
| N exceeding the tolerance | - | - | 2 | 3 |
| **Elekta linac – XiO/Monaco TPS (N=5)** | | | | |
| mean | -0.5% | -0.2% | 0.6% | 1.2% |
| standard deviation | 0.5% | 0.5% | 0.5% | 0.8% |
| N exceeding the tolerance | - | - | - | - |

### 3.3.2 | CT to density conversion

The majority of institution's CT scanners were radiotherapy dedicated. All involved TPS had introduced a customised CT to RED/mass density conversion curve.

The on-site verification of CT to density conversion with SHANE revealed a general failure in trabecular bone (74%) and cortical bone (95%) reference materials. Some other discrepancies were noticed and corrected during the on-site visit, being the H&N IMRT plan calculated using the

updated CT to density curve. In one institution, a difference of about 212 HU was observed in air as the first point of the curve introduced in TPS had a RED of 0.19; a data point corresponding to water was added in another institution as a discrepancy of 44 HU was revealed. A major correction was done in two institutions where the CT to RED curve was built using a combination of CIRS Head & Torso phantom (Model 002H9K) and CATPHAN (Model 600) corresponding reference materials. CATPHAN is not a recommended phantom for this purpose but for CT number constancy check only [35]. The introduced curve on TPS had a shape as shown in Figure 3.1 – "Cal 2016 CIRS Head & Torso + CATPHAN" which had not been noticed by the local medical physics team. Thus differences up to 239 in trabecular bone and 394 HUs in cortical bone reference materials were registered – "Cal 2018 CIRS SHANE". Data points corresponding to CATPHAN materials were then removed and the resulting CT to RED curve – "Cal 2016 CIRS Head & Torso" – was similar to the one obtained on-site with SHANE, as can be seen in Figure 3.1.

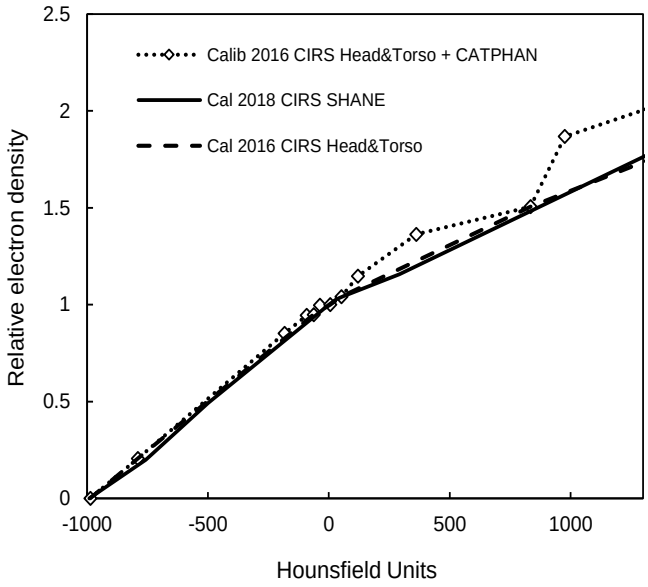


**Figure 3.1** – Comparison between CT to RED conversion curves obtained with data from measurements carried out in 2016 and in the IMRT audit in one of the visited institutions.

### 3.3.3 | Treatment plans

The treatment planning exercise was considered challenging, although all but three created plans fulfilled PTVs coverage and organs-at-risk dose limits. Some characteristics of the H&N treatment plans related to the delivery technique are presented in Table 3.3.

From the 20 H&N IMRT plans, 15 were planned for VMAT, three sliding window, one step & shoot and one helical IMRT. Most institutions performing VMAT (10/15) used two arcs. All sliding window plans had 9 beams. A great difference (about 2 times, on average) in total number of MU between sliding window and VMAT plans was noticed as well as the large spread of total MUs of VMAT plans created in Eclipse. For VMAT plans, the number of control points (CP) per arc was constant (178) and always higher in Eclipse, while the number of CPs was variable in Elekta plans (ranging from 69 to 136). A huge difference was observed between the number of CPs per field for the sliding window plan created in Monaco to be delivered by a Varian linac (53.2) in comparison to plans created in Eclipse, for the same delivery machine type and technique (187.1 on average).

Pre-treatment verification of the created plans was performed as per local routine to evaluate its deliverability and acceptability of the QA results. All plans were considered deliverable and acceptable for treatment.

**Table 3.3** – Treatment plans characteristics.

| | Delivery technique | | |
|---|---|---|---|
| | VMAT (N=15) | Sliding Window (N=3) | Step & Shoot (N=1) |
| **# Fields/Arcs** | 2 Arcs (N=10);<br>3 Arcs (N=2);<br>4 Arcs (N=3) | 9 fields (N=3) | 7 fields (N=1) |
| **Total #MU/plan** | | | |
| Plans Eclipse | | | |
| mean±SD | 667.8±164.3 (N=11) | 1548.0±49.5 (N=2) | - |
| range | 478.3 – 961 | 1513 – 1583 | - |
| Plans Monaco/XiO | | | |
| mean±SD | 775.9±73.4 (N=4)* | 1215 (N=1) | 843.6 (N=1) |
| range | 683.8 – 863.2 | - | - |
| **Total #CP/Field/Arc** | | | |
| Plans Eclipse | | | |
| mean±SD | 178 (N=11) | 187.1±21.2 (N=2) | - |
| range | - | 172.1 – 202.1 | - |
| Plans Monaco/XiO | | | |
| mean±SD | 105±28 (N=4) | 53.2 (N=1) ** | 35.4 (N=1) *** |
| range | 69 – 136 | - | - |

*Elekta VMAT plans had only 1 field that encompassed 2 arcs (3 institutions) or 4 arcs (1 institution); ** Plan created in Monaco to be delivered by a Varian linac; *** 2 CPs per segment.

### 3.3.4 | Output check

In most institutions, 17/20, the daily output of the machine was within the established tolerance of ±2% from reference. The percent difference between calculated – 2 Gy – and measured doses using the audit dosimetry system was on average -0.6%±0.9% (1SD), varying from -2.4% to 0.8%. Beam output on the audit day was taken into account for the subsequent measurements of the SHANE phantom.

### 3.3.5 | IMRT measurements in SHANE phantom

In total, 20 H&N IMRT plans were verified through ionization chamber measurements and EBT3 film irradiation. SHANE phantom was positioned on the treatment couch, and aligned with lasers. Phantom positioning was verified according to the local standard H&N image guided radiation therapy (IGRT) method. Most of the institutions, 17/20 used kV CBCT, 1/20 used MV portal imaging, 1/20 MVCT and in one institution phantom alignment was done based on lasers only, as it was not possible to use the local IGRT method (MV portal imaging) due to technical problems.

Percent differences between the calculated doses corrected for the daily output variation and the ionization chamber measured doses at each reference point are presented in Figure 3.2. Differences were within the established tolerances of ±5% for PTVs and ±7% for spinal cord in all institutions, being below ±3% in most of them. Average differences were -0.6±2.0% (1SD) for the measurement point in PTV_7000, 0.4±1.9% (1SD) in PTVn1_6000, -0.1±2.0% (1SD) in PTVn2_5400 and 0.2±2.2% (1SD) in spinal cord. A major deviation in the point measurement located in spinal cord was registered in one institution. The difference between calculated and measured dose was -12.7%. The high gradient together with the low calculated dose (mean of ~0.5 Gy/fraction) and plan complexity were the identified reasons for the deviation. A new treatment plan was created afterwards by the local team and a follow-up visit was arranged. The difference

was totally resolved, and therefore the initial result was not shown in the graphs and it was excluded from the statistics.

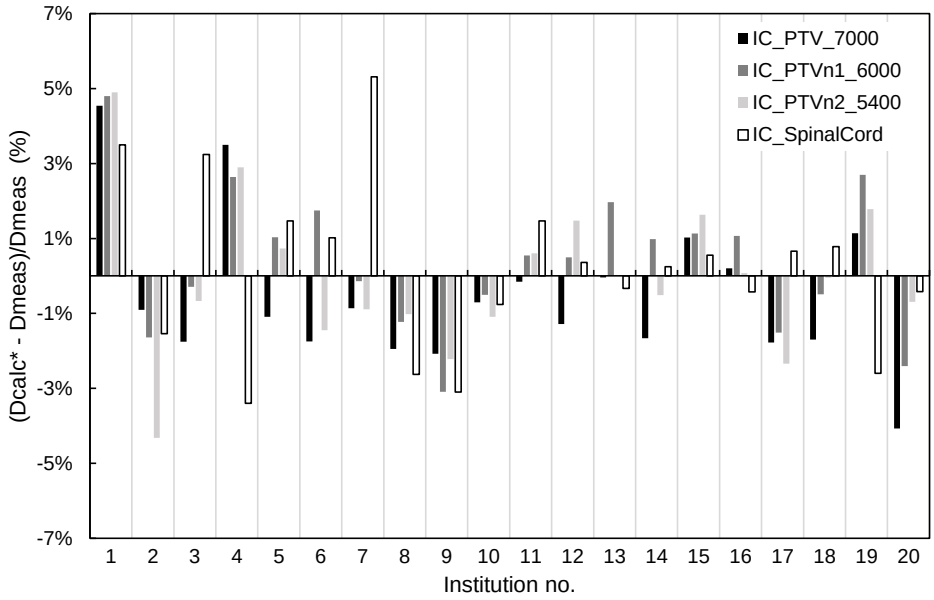


**Figure 3.2 –** Percent dose difference between calculated dose corrected for daily output variation - $D_{calc*}$ - and measured dose in PTVs and spinal cord.

Regarding film analysis, gamma passing rates were on average 96.9±2.9% (1SD), ranging from 90.3% to 99.1%, all above the acceptance limit of 90% for a criterion of 3% global dose/3 mm, 20% threshold. The correlation between ionization chamber measurements and film results was investigated using equation (3.2). As can be seen in Figure 3.3a) there is clearly a group of institutions where higher gamma passing rates (> 95%) correspond to IC measurement results within 3%. When considering only PTVs, the separation between the two groups of institution results is even more evident – Figure 3.3b). The group of institutions with higher gamma passing rates (> 95%) and even lower differences between the IC measured and TPS calculated doses (< 2%) is clearly separated from a group of institutions with poorer film and IC results. There is also a borderline result with a passing rate of 95.1% and IC result >2% located in the upper right quadrant in Figure 3.3b).

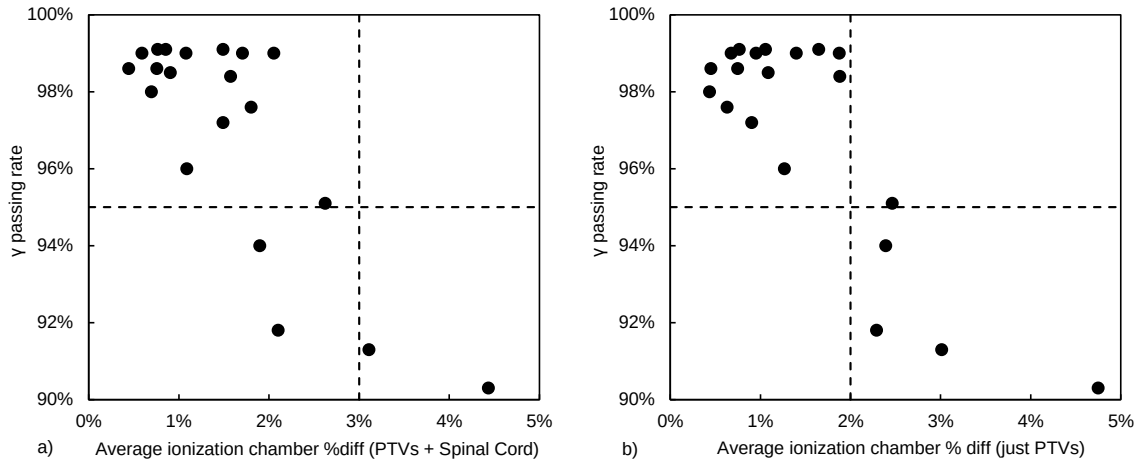


**Figure 3.3 –** Correlation between film gamma passing rates and average absolute ion chamber deviations: a) including all four IC measuring points; b) excluding spinal cord measurement point.

## 3.4 | Discussion

A comprehensive audit program developed by the IAEA to assess the quality of H&N IMRT treatments through on-site visits was implemented in Portugal. To overcome the audit running costs a kick-off workshop was organised with an attractive scientific program and technical exhibition which helped to raise the required funds.

The audit established methodology comprises the dosimetric verification of an H&N IMRT plan created by each participating institution, and a set of tests to evaluate small field dosimetry, MLC performance, the machine beam output and the CT numbers to RED/mass density conversion curve.

The modelling of small fields in TPS is one of most important steps to ensure IMRT dose calculation accuracy [36]. To verify its implementation, output factors were calculated for field sizes ranging from 2×2 $cm^2$ to 6×6 $cm^2$ and compared with reference data published by the IROC-Houston [27,28]. This comparison showed that the TPS tended to overestimate OF, being the deviations larger with decreasing field size. Moreover, differences for the 2×2 $cm^2$ and 3×3 $cm^2$ fields were higher in the Varian linac/Eclipse TPS group than in Elekta linac/XiO/Monaco TPS one. These results were in line with the findings of an IAEA multinational audit of calculated small field output factors that included data from more than 200 different beams [20]. The agreement between inplane and crossplane profiles for a 2×2 $cm^2$ MLC-shaped field calculated in the TPS and the measured ones with EBT3 film was also evaluated. Both field size and penumbra widths were checked and differences were within ±2 mm in all institutions. However, the agreement was less satisfying for the crossplane profiles, which may be explained by the modelling of the leaf ends, TPS commissioning of small fields and MLC positioning uncertainties [21]. Anyway, a very good consistency of data was observed among institutions having the same technology (linac/MLC/TPS).

Another important factor to take in consideration in IMRT is the MLC leaf positioning accuracy as even small deviations can lead to significant dose differences [37,38]. Results of the Picket Fence test showed that the maximum and the median leaf bias were well within the established tolerances of ±1 mm and ±0.5 mm, respectively.

CT to density conversion of either relative electron or mass density is essential to calculate dose in inhomogeneous media being also verified during the on-site visit. A general failure was observed in trabecular and cortical bone reference materials, when considering a tolerance of 20 HU. However, this tolerance is probably too strict. Kilby et al. [39] in 2002 and Nakao et al. [40] more recently, in 2017, established relative electron density acceptance levels for different tissue types. Nakoa et al. proposed a tolerance of ±0.053 in RED difference for 6 MV and ±0.044 for 6FFF MV for bone. These would correspond to around ±100 HU and ±80 HU, for 6 MV and 6FFF MV, respectively. If adopting these tolerances, the percent of failures would decrease to 5% in trabecular bone and 53% in cortical bone. The persisting differences may be due to the use of different calibration phantoms, tissue substitute materials and lack of points in the high density region of the conversion curves. A recommendation on the revision of the materials utilized in that region was given.

Regarding reference dosimetry and the corresponding daily output checks, a very good agreement was obtained with the percent difference between TPS calculated and measured doses of -0.6%±0.9% (1SD), averaged among institutions. Nevertheless, some issues have been noticed that may have contributed for some of the higher deviations. One example is the dose calculation in reference conditions for Monte Carlo based TPSs, which requires a careful assessment namely concerning the adopted calculation dose uncertainty and the grid size.

The dosimetric verification of the created H&N IMRT plan using both ionization chamber and EBT3 film showed that all the audit plans were within the established tolerances. Only four

institutions had passing rates below 95% and IC differences above 2% in PTVs – Figure 3.3b). The greater uncertainty in the IMRT dose delivery for these four institutions may be explained by some identified factors that include:

- Suboptimal plan dose distribution: reference measurement point(s) surrounded by a high dose gradient;
- Treatment plan complexity: much higher number of MUs or lower number of control points than other institutions using the same treatment technique;
- Inclusion of the treatment couch: 3 of these institutions did not account for the couch in treatment planning;
- Phantom positioning verification: alignment according to the lasers only or based just on planar MV imaging;
- Equipment age: audited linac was more than 10 years old in 3 of the 4 institutions; 2 institutions had a quite old TPS version which made it impossible to export the dose distribution in the coronal SHANE phantom corresponding to film with a resolution of about 1 mm, which could have affected the gamma analysis results [41];
- Small beam dosimetry modelling in the TPS: in 2/4 institutions, output factors for 3×3 $cm^2$ and 2×2 $cm^2$ field sizes were out of tolerance;
- Other factors may include experience of working with a newly installed TPS and consideration of all treatment plan parameters and its influence on dose calculation; heavy workload which made it difficult to dedicate enough time to perform the pre-visit activities on time and inherent audit preparation.

These four institutions used different IMRT techniques (SW, SS, and VMAT) and TPSs (Monaco, XiO, Eclipse). Generally no differences between groups of TPSs were identified in terms of deliverability, contrary to what has been reported by Clark et al. [17]. Clark et al., in a multi-institutional audit of rotational IMRT in the UK have shown that the gamma pass rates obtained for measurements performed using the audit system varied between TPS types. Plans created using TPSs designed for the manufacturer's own treatment delivery system (Eclipse and VoLO) gave significantly higher passing rates than the ones designed to be independent of the vendor (Monaco, OMP, and Pinnacle). It must be stressed that the pre-treatment QA verification had not predicted the sub-optimal results in any of these four institutions. Indeed, as other authors have reported, patient-specific QA results were poorly correlated with audit measurements [15,42], failing at identifying institutions that would not meet the acceptability criteria.

The audit plans had associated different levels of complexity as suggested by the total plan MU and the number of control points. VMAT plans created in Eclipse had a constant number of control points, 178, while it was variable in VMAT plans created in Monaco, being here a user definable parameter. Overall, no differences in terms of deliverability (audit results) were identified between these two groups of plans. Regarding the three sliding window plans, they have been calculated in different TPSs but all to be delivered by a Varian linac. The Monaco SW plan had much less control points per field (53.2) than the Eclipse SW plans (172.1 and 202.1) and this plan performed poorly than the Eclipse SW plans, belonging to the group of institutions with suboptimal results. The complexity of the created IMRT/VMAT plans and the correlation with the audit results will be further investigated in a future work.

For film analysis, the methodology proposed by the IAEA was strictly followed to enable comparisons among participating institutions from different countries. The influence of choosing different film dosimetry protocols and following different methodologies on the audit results, namely gamma passing rates, was not highlighted in this paper nor reported as an audit result as it was addressed in a previous work [43].

Portugal was one of the first countries that have implemented this IMRT/VMAT audit at the national level and the results occurred to be more consistent than those of the IAEA multicentre pilot study [26] based on which 5% dose agreement and 90% gamma pass rate criteria were established. The national audit results in Portugal suggest that tighter tolerances could be adopted such as 3% dose difference for ionization chamber measurements and 95% gamma passing rate for film analysis, as 80% (16/20) institutions were able to meet these tighter criteria. To-date there have been 20 countries in Europe that requested a SHANE phantom from the IAEA for 2018-2021 together with the assistance of international experts that facilitate the audit by the national auditing organization. Once more results from different countries become available, the IAEA established pass/fail criteria based on the multicentre pilot may be re-examined.

## 3.5 | Conclusion

The IAEA supported national audit methodology for IMRT dose delivery verification has been carried out in Portugal in 2018 and covered 100% of radiotherapy institutions performing IMRT treatments.

Analysis of tests to check basic small field dosimetry data and MLC performance showed a good compliance with the established tolerances. Verification of the created H&N IMRT plans revealed a general good agreement between calculated and measured dose distributions. Overall, the audit results showed that the status of TPS calculations and delivery for H&N IMRT in Portugal are within the specified tolerances. At the same time the audit identified factors that contributed to increased uncertainties in the IMRT dose delivery in some institutions, and the relevant recommendations for quality improvement were made.

This initiative was certainly important to increase the confidence of all involved professionals and to strengthen the cooperation among the Portuguese medical physics community, giving place for future scientific projects.

## 3.6 | Funding

Part of this work was supported by FCT, the Portuguese Foundation for Science and Technology, through a PhD scholarship, reference SFRH/BD/118929/2016.

## 3.7 | 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.

## 3.8 | Acknowledgments

The authors would like to thank staff from all audit participating institutions; Pavel Kazantsev, Paulina Wesolowska, Miguel Capela, Tiago Ventura and Josefina Mateus for their expert advice; the IAEA for lending the phantom; IPOCFG, E.P.E, for acting as pilot centre; Medical Physics Division of the Portuguese Physics Society for supporting the project.

## 3.9 | References


[1] Nutting CM, Morden JP, Harrington KJ, Urbano TG, Bhide SA, Clark C, et al. Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): a phase 3 multicentre randomised controlled trial. Lancet Oncol 2011;12:127–36. https://doi.org/10.1016/S1470-2045(10)70290-4.

[2] Izewska J, Lechner W, Wesolowska P. Global availability of dosimetry audits in radiotherapy: The IAEA dosimetry audit networks database. Phys Imaging Radiat Oncol 2018;5:1–4. https://doi.org/10.1016/j.phro.2017.12.002.

[3] Ibbott GS, Thwaites DI. Audits for advanced treatment dosimetry. J Phys Conf Ser 2015;573:12002. https://doi.org/10.1088/1742-6596/573/1/012002.

[4] Clark CH, Jornet N, Muren LP. The role of dosimetry audit in achieving high quality radiotherapy. Phys Imaging Radiat Oncol 2018;5:85–7. https://doi.org/10.1016/j.phro.2018.03.009.

[5] Clark CH, Aird EGA, Bolton S, Miles EA, Nisbet A, Snaith JAD, et al. Radiotherapy dosimetry audit: three decades of improving standards and accuracy in UK clinical practice and trials. Br J Radiol 2015;88:20150251. https://doi.org/10.1259/bjr.20150251.

[6] Weber DC, Vallet V, Molineu A, Melidis C, Teglas V, Naudy S, et al. IMRT credentialing for prospective trials using institutional virtual phantoms: results of a joint European Organization for the Research and Treatment of Cancer and Radiological Physics Center project. Radiat Oncol 2014;9:123. https://doi.org/10.1186/1748-717X-9-123.

[7] Miri N, Lehmann J, Legge K, Zwan BJ, Vial P, Greer PB. Remote dosimetric auditing for intensity modulated radiotherapy: A pilot study. Phys Imaging Radiat Oncol 2017;4:26–31. https://doi.org/10.1016/j.phro.2017.11.004.

[8] Nakamura M, Minemura T, Ishikura S, Nishio T, Narita Y, Nishimura Y. An on-site audit system for dosimetry credentialing of intensity-modulated radiotherapy in Japanese Clinical Oncology Group (JCOG) clinical trials. Phys Med 2016;32:987–91. http://doi.org/10.1016/j.ejmp.2016.07.002.

[9] Izewska J, Andreo P. The IAEA/WHO TLD postal programme for radiotherapy hospitals. Radiother Oncol 2000;54:65–72. https://doi.org/10.1016/S0167-8140(99)00164-4.

[10] Gershkevitsh E, Schmidt R, Velez G, Miller D, Korf E, Yip F, et al. Dosimetric verification of radiotherapy treatment planning systems: results of IAEA pilot study. Radiother Oncol 2008;89:338–46. https://doi.org/10.1016/j.radonc.2008.07.007.

[11] Gershkevitsh E, Pesznyak C, Petrovic B, Grezdo J, Chelminski K, Lopes MC, et al. Dosimetric inter-institutional comparison in European radiotherapy centres: Results of IAEA supported treatment planning system audit. Acta Oncol 2014;53:628–36. https://doi.org/10.3109/0284186X.2013.840742.

[12] Lopes MC, Cavaco A, Jacob K, Madureira L, Germano S, Faustino S, et al. Treatment planning systems dosimetry auditing project in Portugal. Phys Medica 2014;30:96–103. https://doi.org/10.1016/j.ejmp.2013.03.008.

[13] Lafond C, Chiavassa S, Bertaut C, Boussion N, Chapel N, Chapron L, et al. DEMAT: A multi-institutional dosimetry audit of rotational and static intensity-modulated radiotherapy. Phys Medica 2019;32:664–70. https://doi.org/10.1016/j.ejmp.2016.04.008.

[14] Seravalli E, Houweling AC, Battum L Van, Raaben TA, Kuik M, de Pooter JA, et al. Auditing local methods for quality assurance in radiotherapy using the same set of predefined treatment plans. Phys Imaging Radiat Oncol 2018;5:19–25. https://doi.org/10.1016/j.phro.2018.01.002.

[15] Jurado-Bruggeman D, Hernández V, Sáez J, Navarro D, Pino F, Martínez T, et al. Multi-centre audit of VMAT planning and pre-treatment verification. Radiother Oncol 2017;124:302–10. https://doi.org/10.1016/j.radonc.2017.05.019.

[16] Clark CH, Hansen VN, Chantler H, Edwards C, James H V, Webster G, et al. Dosimetry audit for a multi-centre IMRT head and neck trial. Radiother Oncol 2009;93:102–8. https://doi.org/10.1016/j.radonc.2009.04.025.

[17] Clark CH, Hussein M, Tsang Y, Thomas R, Wilkinson D, Bass G, et al. A multi-institutional dosimetry audit of rotational intensity-modulated radiotherapy. Radiother Oncol 2014;113:272–8. http://dx.doi.org/10.1016/j.radonc.2014.11.015.

[18] Veres, A., Hallet JX. Equal-Estro experience in dosimetry audits in advanced techniques of radiotherapy – the tomotherapy example. SSDL Newsletter No. 70 2019:29–31.

https://www-pub.iaea.org/MTCD/Publications/PDF/Newsletters/ssdl-70.pdf [accessed 26 June 2019].

[19] Molineu A, Hernandez N, Nguyen T, Ibbott G, Followill D. Credentialing results from IMRT irradiations of an anthropomorphic head and neck phantom. Med Phys 2013;40. https://doi.org/10.1118/1.4773309.

[20] Lechner W, Wesolowska P, Azangwe G, Arib M, Gabriel V, Alves L, et al. A multinational audit of small field output factors calculated by treatment planning systems used in radiotherapy. Phys Imaging Radiat Oncol 2018;5:58–63. https://doi.org/10.1016/j.phro.2018.02.005.

[21] Izewska J, Wesolowska P, Azangwe G, Followill DS, Thwaites DI, Arib M, et al. Testing the methodology for dosimetry audit of heterogeneity corrections and small MLC-shaped fields: Results of IAEA multi-center studies. Acta Oncol 2016;55:909–16. https://doi.org/10.3109/0284186X.2016.1139180.

[22] Carson ME, Molineu A, Taylor PA, Followill DS, Stingo FC, Kry SF. Examining credentialing criteria and poor performance indicators for IROC Houston's anthropomorphic head and neck phantom. Med Phys 2016;43:6491. https://doi.org/10.1118/1.4967344.

[23] Izewska J. Development of Quality Audits for Advanced Technology in Radiotherapy Dose Delivery. SSDL Newsletter No 64 2016;8–15. https://www.iaea.org/publications/11032/ssdl-newsletter-issue-no-64-february-2016 [accessed 26 June 2019].

[24] International Atomic Energy Agency (IAEA). WORKING MATERIAL IAEA Supported National "End -to- End" Audit Programme for Dose Delivery Using Intensity Modulated Radiation Therapy through On-Site Visits to Radiation Therapy Institutions, 2019.https://dosimetry-audit-networks.iaea.org/Content/end-to-end%20CIRS%20SHANE/National%20end-to-end%20IMRT-VMAT%20audit%20methodology.pdf [accessed 13 June 2019].

[25] CIRS. Phantom Patient for VMAT & IMRT.

http://www.cirsinc.com/products/all/119/shane-phantom-patient-for-vmat-and-imrt/ [accessed 26 June 2019].

[26] Kazantsev P, Clark C, Venencia D, Georg D, Gershkevitsh E, Van Dyk, J, et al. New IAEA end-to-end on-site IMRT audit methodology: Pilot test results. Proceedings of the International Conference on Advance in Radiation Oncology (ICARO2). Vienna: International Atomic Energy Agency; 2017.

[27] Followill DS, Kry SF, Qin L, Leif J, Molineu A, Alvarez P, et al. The Radiological Physics Center's standard dataset for small field size output factors. J Appl Clin Med Phys 2012;13:282–9. https://doi.org/10.1120/jacmp.v13i5.3962.

[28] Followill DS, Kry SF, Qin L, Leif J, Molineu A, Alvarez P, et al. Erratum : "The Radiological Physics Center's standard dataset for small field size output factors". J Appl Clin Med Phys 2014;15:356–7. https://doi.org/10.1120/jacmp.v15i2.4757.

[29] International Atomic Energy Agency (IAEA). IAEA TRS-483: Dosimetry of Small Static Fields Used in External Beam Radiotherapy: An International Code of Practice for Reference and Relative Dose Determination. Vienna: International Atomic Energy Agency; 2017.

[30] International Atomic Energy Agency (IAEA). IAEA TRS-430: Commissioning and Quality Assurance of Computerized Planning Systems for Radiation Treatment of Cancer. Vienna: International Atomic Energy Agency; 2005.

[31] Kazantsev P, Hernandez V, Luketin L, Izewska J. EP-2194: Analysis of treatment planning feasibility of the multinational end-to-end IMRT audit methodology. Radiother Oncol 2018;127:S1212–3. https://doi.org/10.1016/S0167-8140(18)32503-9.

[32] International Atomic Energy Agency (IAEA). Absorbed Dose Determination in External Beam Radiotherapy: An International Code of Practice for Dosimetry Based on Standards of Absorbed Dose to Water. IAEA TRS-398. Vienna: International Atomic Energy Agency; 2000.

[33] Lewis D, Micke A, Yu X, Chan MF. An efficient protocol for radiochromic film dosimetry combining calibration and measurement in a single scan. Med Phys 2012;39:6339–50. https://doi.org/10.1118/1.4754797.

[34] Klein EE, Hanley J, Bayouth J, Carolina N, Simon W, Dresser S, et al. Task Group 142 report: Quality assurance of medical accelerators a. Med Phys 2009; 36:4197–212. https://doi.org/10.1118/1.3190392.

[35] Husby E, Svendsen ED, Andersen HK, Martinsen ACT. 100 days with scans of the same Catphan phantom on the same CT scanner. J Appl Clin Med Phys 2017;18:224–31. https://doi.org/10.1002/acm2.12186.

[36] ICRU Report 91: Prescribing, Recording, and Reporting of Stereotactic Treatments with Small Photon Beams. J ICRU, 2014;14:31-53.

[37] Chang CC, Lin JC, Cheng HW, Tsai JT. Dose Impact of Systemic MLC Position Error for Esophagus Cancer Plan: Volumetric Modulated Arc Therapy Versus Step and Shoot Modulated Therapy. Int J Radiat Oncol Biol Phys 2015;93:E563. https://doi.org/10.1016/j.ijrobp.2015.07.1989.

[38] Mu G, Ludlum E, Xia P. Impact of MLC leaf position errors on simple and complex IMRT plans for head and neck cancer. Phys Med Biol 2008;53:77–88. https://doi.org/10.1088/0031-9155/53/1/005.

[39] Kilby W, Sage J, Rabett V. Tolerance levels for quality assurance of electron density values generated from CT in radiotherapy. Phys Med Biol 2002;47:1485–92. https://doi.org/10.1088/0031-9155/47/9/304.

[40] Nakao M, Ozawa S, Yamada K, Yogo K, Hosono F, Hayata M, et al. Tolerance levels of CT number to electron density table for photon beam in radiotherapy treatment planning system. J Appl Clin Med Phys 2018;19:271–5. https://doi.org/10.1002/acm2.12226.

[41] Huang JY, Pulliam KB, McKenzie EM, Followill DS KS. Effects of spatial resolution and noise on gamma analysis for IMRT QA. J Appl Clin Med Phys 2014;15:4690. https://doi.org/10.1120/jacmp.v15i4.4690.

[42] Kry SF, Molineu A, Kerns J, Faught A, Huang JY, Pulliam K, et al. Institutional patient-specific intensity-modulated radiation therapy quality assurance does not predict unacceptable plan delivery as measured by IROC Houston's head and neck phantom. Int J Radiat Oncol Biol Phys 2015;90:1195–201. https://doi.org/10.1016/j.ijrobp.2014.08.334.

[43] Santos T, Ventura T, Capela M, Mateus J, Lopes MC. Influence of Film Dosimetry Protocols on IMRT Audit Results. Proceedings of the International Symposium on Standards, Applications and Quality Assurance in Medical Radiation Dosimetry (IDOS2019). Vienna: International Atomic Energy Agency; 2019.

## 3.10 | Supplementary materials

**Table S.3.1** – Dose-volume constraints for the H&N plan.

| Structure | Volume | Dose | Planning priority* |
|---|---|---|---|
| PTV_7000 | 98% | >90% (63.0 Gy) | 2 |
| (primary nasopharynx) | 95% | >95% (66.5 Gy) | |
| | 50% | =100% (70.0Gy) | |
| | 2% | <107% (74.9 Gy) | |
| PTVn1_6000 | 98% | >90% (54.0Gy) | 3 |
| (involved nodes) | 95% | >95% (57.0Gy) | |
| | 50% | 60.0-62.0 Gy | |
| PTVn2_5400 | 98% | >90% (48.6Gy) | 4 |
| (elective nodes) | 95% | >95% (51.3Gy) | |
| | 50% | 54.0-56.0 Gy | |
| SpinalCord | 2% | <45Gy | 1 |
| SpinalCord_03 | 2% | <50Gy | |
| BrainStem | 2% | <50Gy | 1 |
| BrainStem _03 | 2% | <55Gy | |
| Parotid_L | Mean | <24Gy | 5 |
| Parotid_R | Mean | as low as possible | 6 |

*Planning priority: 1-highest, 6-lowest

## 3.11 | Adaptation of the audit methodology to tomotherapy

IMRT treatments are performed at the audit pilot centre, IPOCFG, using helical IMRT in a Tomotharapy HD unit (Accuray, Sunnyvale, CA, USA). As the IAEA designed methodology does not contemplate helical tomotherapy (HT), some adaptations had to be considered, which may be useful for future developments. A set of dedicated tests to check machine beam output, the beam model and small beam dosimetry were defined. The H&N plan verification is independent of the IMRT technique, therefore no modifications were needed. The MLC performance was also not directly assessed.

Helical Tomotherapy is a pre-commissioned machine in the sense that it comes to the installation site with a gold standard beam model that only needs to be verified during the commissioning phase [1]. In HT, the reference dosimetry conditions established in conventional codes of practice (CoP) [2,3] do not apply, as the 10×10 $cm^2$ homogeneous field cannot be configured. The IAEA/AAPM international CoP TRS 483, issued late in 2017, extends reference dosimetry to small and non-reference fields by introducing the concept of machine-specific reference (msr) field. The reference dosimetry setup in the HT unit corresponds to a 5×10 $cm^2$ static field size defined at a source-to-skin distance of 85 cm, and depth of 10 cm [4].

By the time the HT machine was installed at IPOCFG, in 2016, the TRS 483 had not yet been released and the Accuray acceptance tests do not contemplate the concept of reference dosimetry. Instead, the machine absolute dose calibration is verified by considering a helical IMRT reference plan created in the TPS to deliver a uniform dose to a cylindrical structure in the centre of the Tomotherapy phantom (Cheese Phantom), which constitutes a plan-class specific reference (pcsr) field [5]. TG-148 [1] also recommends the use of a plan like this for rotational output verification.

After the launch of TRS 483, the given recommendations were followed at IPOCFG in order to determine dose in reference conditions in HT, using the standard dosimetry system composed by an Exradin A1SL ionization chamber and a TomoElectrometer (Standard Imaging, Middleton, WI, USA) – please see Appendix I. Still, as it is not possible to calculate dose under reference conditions in the TPS, the consistency of the machine output with the TPS can only be verified in rotational mode [1].

In the context of the IMRT audit, the machine beam output verification was performed in two steps, which can be adopted by any other institution. First, to verify the beam output consistency with the TPS, the helical IMRT plan was delivered and the local HT standard dosimetry system was used for dose determination, as the auditor's system did not fit in the Cheese phantom. Second, the dose determined using the local and auditor's dosimetry systems was compared in the machine-specific reference set-up (field size of 5×10 $cm^2$, depth of 10 cm and SSD of 85 cm). The difference between the dosimetric systems was below 1%, therefore, the relative output of the day was taken from the measured dose in rotational mode.

To check the beam model, an independent verification may be performed. At IPOCFG, longitudinal and transverse dose profiles for the three field sizes 1, 2.5 and 5 cm lengths were measured and compared to the gold standard. That was done using a MP3-T motorized water phantom from PTW and two ionization chambers, the Exradin A1SL and a PTW 31016 PinPioint 3D chamber – Appendix II. For the purpose of the audit, the same procedure can be adopted utilizing radiochromic films placed in a water equivalent phantom instead, being the field size and penumbras compared to the gold standard beam data, similarly to what has been done for the 2×2 $cm^2$ field in the conventional linacs.

The information conveyed in TRS 483 was also explored at IPOCFG to find a sustained way of performing relative dosimetry, namely the determination of output factors in HT – please see Appendix I. If there were HT machines to be audited in sufficient number to establish a consistent

dataset, the measured output factors with radiochromic film, for instance, could then be compared with the reference values.

All the described procedures to check the machine beam output, the beam model and small field dosimetry could eventually be considered to extend the audit methodology to HT.

# References

[1] Langen KM, Papanikolaou N, Balog J, et al. QA for helical tomotherapy: Report of the AAPM Task Group 148. Med Phys. 2010;37(9):4817-4853. doi: https://doi.org/10.1118/1.3462971

[2] International Atomic Energy Agency (IAEA). Absorbed Dose Determination in External Beam Radiotherapy: An International Code of Practice for Dosimetry Based on Standards of Absorbed Dose to Water. IAEA TRS-398. Vienna: International Atomic Energy Agency; 2000

[3] Almond PR, Biggs PJ, Coursey BM, et al. AAPM's TG-51 protocol for clinical reference dosimetry of high-energy photon and electron beams. Med Phys. 1999;26(9):1847-1870. doi: https://doi.org/10.1118/1.598691

[4] International Atomic Energy Agency (IAEA). Dosimetry of small static fields used in external beam radiotherapy. Technical Reports Series No. 483. Vienna: International Atomic Energy Agency; 2017

[5] Alfonso R, Andreo P, Capote R, et al. A new formalism for reference dosimetry of small and nonstandard fields. Med Phys. 2008;35(11):5179-5186. doi: https://doi.org/10.1118/1.3005481

# CHAPTER 4

# Evaluation of the complexity of treatment plans from a national IMRT/VMAT audit – Towards a plan complexity score

Physica Medica, 2020, volume 70, pages 75-84

T Santos[a,b], T Ventura[b], MC Lopes[b]

[a] Physics Department, University of Coimbra, Coimbra, Portugal
[b] Medical Physics Department, IPOCFG, E.P.E., Coimbra, Portugal

## Abstract

**Purpose:** This work aimed to characterize and compare the complexity of the plans created in the context of a national IMRT/VMAT audit. A plan complexity score is proposed to summarize all the evaluated complexity features.

**Materials and Methods:** Nine complexity metrics have been computed for the audit plans, evaluating different complexity aspects. An approach based on Principal Component Analysis was followed to explore the correlation between the metrics and derive a smaller set of new uncorrelated variables (principal components, PCs). The resulting PCs were then used to calculate a plan complexity score. Plan quality was also assessed and the correlation between plan complexity, quality and deliverability investigated using the Spearman's rank correlation coefficient.

**Results:** The first two PCs explained over 90% of the total variance in the original dataset. Their representation allowed to identify patterns in the data, namely a clear separation between plans created using different technologies/techniques. The calculated plan complexity score quantified these differences. Sliding window Eclipse plans were found to be the most complex and VMAT Eclipse group presented the highest complexity variability, for the evaluated parameters. Concerning plan quality, no differences between treatment technology/technique have been identified. However, plans with larger number of monitor units tended to be associated with higher deviations between calculated and measured doses.

**Conclusions:** The proposed plan complexity score allowed to summarize the differences not only inter- but also intra-groups of technologies/techniques, paving the way for improvement of the planning strategies at the national level through knowledge sharing.

**Keywords:** IMRT/VMAT audit, plan complexity, principal component analysis, plan complexity score

## 4.1 | Introduction

Intensity modulated radiation therapy (IMRT), including volumetric modulated arc therapy (VMAT), has become extensively used to treat cancer patients. However, the known benefits [1,2] come along with an increased potential for errors. Recognizing the effectiveness of dosimetry audits in identifying discrepancies and reducing uncertainties, the International Atomic Energy Agency (IAEA) developed an 'end-to-end' audit program to review the physics aspects of Head and Neck (H&N) IMRT/VMAT treatments through on-site visits. Portugal was one of the first countries implementing this audit at a national level. The national experience, results and findings have been recently published [3]. According to the established methodology, all participating institutions had to create a treatment plan, based on a common list of planning objectives, which was verified afterwards through on-site measurements. However, due to the numerous degrees of freedom existing in IMRT, the created plans could have associated different levels of complexity.

Complexity has been defined as "the frequency and amplitude of fluctuations in intensity distribution of a beam" [4]. A high complexity level is generally characterized by a large number of monitor units (MU), and small, narrow, off-axis and/or irregularly shaped apertures [5]. As reported, the degree of complexity of a beam/plan can significantly compromise treatment deliverability, due to multileaf collimator (MLC) positioning accuracy, linear accelerator (linac) performance and/or limitations in dose calculation [5,6]. Therefore, when comparing plans with a similar dose distribution, less complex ones are typically preferred as the associated uncertainties tend to be lower. Having this in mind, many authors have developed and adapted different metrics to assess the complexity of static gantry IMRT and VMAT plans [5–12]. Fluence map-based metrics, such as the Fluence Map Complexity (FMC) [7] and the Modulation Index (MI) [8], were the first proposed and both evaluate the variations between neighbour bixel intensities in the fluence maps. Aperture-based metrics like the Modulation Complexity Score (MCS) [6], directly evaluate the shape, size, and/or location of individual beam apertures. Parameters such as the leaf travel, and variations of the dose rate and gantry speed in VMAT treatments have also been assessed [11,12]. Two review papers on complexity metrics have been recently published by Chiavassa et al [13] and Antoine et al [14].

A complexity metric that could a priori identify beams/plans prone to fail pre-treatment quality assurance (QA) measurements would be a useful tool regarding the IMRT planning optimization and verification process. However, despite the strong efforts being made, the reported correlations, when found, depend on the institution's technology, pre-treatment QA verification devices, evaluation metrics and acceptability criteria [5,6,12]. Still, complexity metrics play an important role in treatment plans characterization, giving important insights into the planning and optimization processes, and may also contribute to establish a trade-off between plan quality and complexity [13,15,16]. As their definition is typically independent of the treatment technology/technique, such metrics can be utilized to compare various complexity aspects associated with plans, which may be useful in the context of multi-institutional dosimetry audits. In dosimetry audits, a particular scenario is faced, as the same treatment cases and common deliverability evaluation tools and criteria are typically involved. Therefore, the complexity and quality of the treatment plans can be evaluated and compared in an independent way. Results of the analysis may help to benchmark institutions with similar equipment and drive the medical physicists' attention to parameters that contribute to over-modulated plans [15]. There are still few publications reporting results from dosimetry audits [15–18]. Hernandez et al [15], for instance, using 40 plans from a VMAT audit and 100 clinical plans, have computed multiple complexity metrics proposed in the literature and examined the relationship between them. They found that if

some indices are strongly correlated, others may lead to a distinct solution when trying to rank the plans according to their complexity, concluding that it is not trivial to choose a single metric. Indeed, there is not yet a single metric that can incorporate all the complexity aspects of a plan, which makes difficult the assessment of plan complexity.

A promising approach to simplify the complexity analysis is Principal Component Analysis (PCA). PCA is a multivariate technique that is broadly used [19–21] to reduce the dimensionality of datasets by exploring the correlation between variables, while preserving most of the initial information. The dimensionality reduction eliminates redundancy and typically helps to summarize, visualize and identify patterns in the data [22].

This study builds on a previous publication [3] through further analysis of the H&N IMRT/VMAT audit plans data and inclusion of complexity metrics. Plan quality was also assessed using an in-house software tool [23]. A novel approach based on PCA was followed to characterize and compare the complexity of the audit plans, created using different technologies and IMRT/VMAT techniques. Multiple complexity metrics were computed and PCA was employed to reduce the dimensionality of the original dataset. Based on the new uncorrelated variables, we propose a plan complexity score to evaluate the complexity of the audit treatment plans through a single indicator that contains all the features embedded in the multiple metrics.

## 4.2 | Materials and Methods

### 4.2.1 | Treatment plans and plan deliverability evaluation

Nineteen treatment plans were considered in this study resulting from a national IMRT/VMAT audit carried out in Portugal with the IAEA assistance [3]. As per the audit methodology [24], an H&N phantom named Shoulder, Head and Neck End-to-End (SHANE – Computerized Imaging Reference Systems (CIRS) Inc., Norfolk, Virginia, USA) and a set of pre-delineated structures were used to simulate a clinical case of a patient with a nasopharynx tumour. Each institution had to create a treatment plan considering its IMRT standard planning technique and the usual dose calculation settings, taking into account the provided dose prescriptions and constraints. The audited treatment planning systems (TPS) were: Eclipse (n=13), Monaco (n=5), XiO (n=1), and VoLO (n=1). The IMRT techniques included: VMAT (n=15), sliding window (SW, n=3), step & shoot (SS, n=1) and helical IMRT (n=1). Regarding the treatment machines: Varian linacs (n=14), Elekta linacs (n=5) and Tomotherapy HD (n=1). Helical IMRT was excluded from this work.

Plan deliverability, i.e. the agreement between planned and delivered doses, was evaluated using the local pre-treatment verification methods, followed by independent audit measurements in the SHANE phantom. A summary of the used institution's pre-treatment verification systems and more details about the audit methodology are presented in a previously published paper by Santos et al [3]. The audit measurements were performed using a dedicated dosimetry set, including an ionization chamber (IC) that was placed at four reference locations (3 planning target volumes – PTVs, and spinal cord) inside the SHANE phantom and a Gafchromic EBT3 film (Ashland Inc., Covington, Kentucky, USA) to assess the dose distribution in a coronal plane. FilmQA Pro software (Ashland Inc., Covington, Kentucky, USA) was used and global gamma analysis performed, considering a criterion of 3% dose difference/3 mm distance-to-agreement and 20% threshold (TH). The audit established tolerances for IC point dose measurements were ±5% for PTVs and ±7% for spinal cord, and a gamma passing rate of 90% for film analysis, defining the criteria for deliverability. For the purpose of the present study a stricter criteria of 2%/2 mm 20% TH was also considered for gamma analysis.

### 4.2.2 | Plan complexity metrics

Several metrics that evaluate different complexity features of both static/dynamic IMRT and VMAT plans have been computed from the DICOM RT files of the audit plans using an in-house developed MATLAB R2017b program (Mathworks, Natick, Massachusetts, USA). Based on the information associated to each control point, the following metrics were calculated:

a. Total plan MU [5]. Generally a larger number of MU is related to higher complexity as it is associated with smaller apertures and contributes to increased treatment time.
b. Modulation Complexity Score (MCS) [6]. MCS evaluates the plan modulation by considering the relative variability of the shape and area of the segments, and the corresponding weight (the fraction of total MU delivered). MCS varies between 0 and 1, and when modulation increases, MCS decreases, so we have used 1-MCS in order to harmonize the monotonicity with the other indices.
c. Plan Modulation (PM) [9]. PM characterizes the area of the segments relative to the union area of all segments in the beam/arc. In a plan with many small apertures that are far apart, PM tends to 1.
d. Small Aperture Score (SAS) [5]. SAS quantifies the fraction of MLC leaf pairs with openings less than a given value, 10 mm (SAS10) and 20 mm (SAS20), in this study.
e. Cross-Axis Score (CAS) [5]. CAS assesses the amount of off-axis apertures by determining the fraction of MLC leaf pairs that cross the central beam axis.
f. Mean Asymmetry Distance (MAD) [5]. MAD describes the apertures asymmetry as the weighted average of the lateral distance between the centre of every open MLC leaf pair gap and the central beam axis.
g. Plan Irregularity (PI) [9]. PI gives an indication of the apertures irregularity or narrowness by evaluating their non-circularity. The PI value becomes higher with increasing irregularity/narrowness.
h. Edge Metric (EM) [10]. EM evaluates the complexity of the MLC apertures by calculating the ratio between the length of the MLC leaf sides and the aperture area. High EM values mean that the difference between adjacent leaves position is large.

The dependencies between the computed complexity metrics were investigated through the calculation in MATLAB of the Spearman's correlation coefficients, $r_s$. The correlation between each metric and the audit measurement results was also evaluated. Depending on the absolute value of $r_s$, the strength of a correlation was classified as: 0-0.19 "very weak", 0.20-0.39 "weak", 0.40-0.59 "moderate", 0.60-0.79 "strong" and 0.80-1 "very strong". To account for multiple-testing, Bonferroni correction was applied to get the adjusted p-value, defining the level of statistical significance, initially set at 5%.

### 4.2.3 | Plan complexity score using PCA

PCA is a multivariate technique used to reduce the dimensionality of large datasets by finding a new set of uncorrelated variables (principal components, PCs) that result from a linear combination of the original ones, and that successively maximize variance. The resulting number of PCs is equal to the number of original variables. To decide how many PCs are necessary, a common approach consists in considering a predefined cut-off of the cumulative percentage of the total variance [22].

In this study, PCA was performed in MATLAB, to simplify the complexity analysis by reducing the dimensionality of the original dataset, where each complexity metric represents one dimension. All computed metrics increased with increasing complexity, and were considered in the PCA analysis. This corresponded to a data matrix X with values $x_{ij}$ of 9 metrics for the 19 plans.

Each column was a vector $x_j$ of 19 $i$ values on the *jth* complexity metric. As some indices had different units of measurement and numerical ranges, each data value $x_{ij}$ was standardized, as follows:

$$z_{ij} = \frac{x_{ij} - \overline{x_j}}{s_j} \tag{4.1}$$

where $\overline{x_j}$ is the mean of the 19 values of the *jth* metric and $s_j$ the associated standard deviation. Therefore, X was replaced by the corresponding standardized data matrix Z with values $z_{ij}$.

The output of the PCA analysis consisted of 9 principal components – $PC_k$, (k=1,…,9):

$$PC_k = w_{1j}Z_1 + w_{2j}Z_2 + \cdots + w_{9j}Z_j \tag{4.2}$$

where $w_{lj}$ (l=1,…,9; j=1,…9), also known as loadings, are the PCs coefficients and represent the weight of each original variable j for the principal component. The PC coefficients can range from -1 to 1, being the sum of their square values equal to 1. The higher the absolute value of a PC coefficient, the more the associated variable contributes for the PC [22].

The resulting PCs are uncorrelated as they basically represent orthogonal directions in the original multi-dimensional space that maximize the variance, reflecting different dimensions of the data. This means that the sum of the cross products of the vector formed by the loadings of a given PC with any of the vectors representing the other PCs is equal to zero. The PCs are presented in a descending way such that the first PC corresponds the direction of maximum variability. To decide how many PCs to retain, a cut-off of 80% of the total variance explained was adopted [22].

Based on the PCA outcome, a plan complexity score (PCS) is proposed to characterize and compare the audit plans through a single indicator. The PCS was computed as the weighted mean of the selected principal components:

$$PCS = \sum_{k=1}^{n} \frac{v_k}{v} \times PC_k \tag{4.3}$$

where $n$ is the minimum number of PCs corresponding to a cumulative percentage of the total variance explained higher than 80%. $v$ is the total variance explained by the retained PCs and $v_k$ the percent of variance explained by $PC_k$.

This index quantifies the complexity of a plan in a linear scale and it can have both positive and negative values. The higher the value of PCS, the greater the plan complexity. However, its absolute value may be difficult to interpret. To overcome this, a normalized version of the plan complexity score, $nPCS$, was calculated for a given plan $i$, as:

$$nPCS_i = \frac{PCS_i - min\ PCS}{max\ PCS - min\ PCS} \tag{4.4}$$

According to this definition, nPCS varies between 0 and 1. nPCS is 0 for a plan having the minimum PCS in the studied set (minPCS) and 1 for the plan having the maximum PCS (maxPCS). The higher the value of nPCS, the greater the plan complexity for the group of plans considered in the study.

### 4.2.4 | Plan quality

Plan quality of the plans was assessed and compared using a software tool called SPIDERplan [23]. This tool allows to obtain a global plan score based on the accomplishment of the individual PTVs and organs-at-risk (OAR) dosimetric goals. SPIDERplan was customized for the audit purpose as: 1) all structures were divided in three groups – Critical, PTV and SalivaryGlands; 2) a relative weight was assigned to each group and structure, according to the audit planning priorities and in line with RTOG 0615 criteria [25] – Table 4.1; 3) a score (scalar) was determined for each structure as a ratio between the tolerance dose and the planned dose; and finally, 4) a global plan score was calculated as the weighted sum of the score of each structure for all groups:

$$Global\ plan\ score = \sum_i w_{group(i)} \sum_j w_{struct(j)} score_{struct(j)} \tag{4.5}$$

where $w_{group(i)}$ and $w_{struct(j)}$ are the relative weights assigned to each group i and structure j, respectively. The scores associated to the PTVs (scorePTV) and to the OARs (scoreOAR) were given by:

$$score_{PTV} = \frac{D_{TC,PTV}}{D_{P,PTV}} \tag{4.6}$$

where $D_{TC,PTV}$ is the prescribed dose for the PTV and $D_{P,PTV}$ is the planned dose;

$$score_{OAR} = \frac{D_{P,OAR}}{D_{TC,OAR}} \tag{4.7}$$

where $D_{P,OAR}$ is the planned dose for the OAR and $D_{TC,OAR}$ is the tolerance dose. TC stands for tolerance criterion.

**Table 4.1 –** Groups and structures considered for plan quality evaluation using SPIDERplan. The assigned relative weights are presented within parentheses.

| **Groups** | **Structures** | |
|---|---|---|
| **Name / $w_{group\ (i)}$** | **Name / $w_{struct\ (j)}$** | **Tolerance dose** |
| Critical group (50%) | SpinalCord (50%) | $D_{2\%} \leq 50$ Gy |
| | Brainstem (50%) | $D_{2\%} \leq 55$ Gy |
| PTV group (40%) | PTV_7000 (44.5%) | $D_{98\%} \geq 95\%$* of 70 Gy |
| | PTVn1_6000 (33.3%) | $D_{98\%} \geq 95\%$* of 60 Gy |
| | PTVn2_5400 (22.2%) | $D_{98\%} \geq 95\%$* of 54 Gy |
| SalivaryGlands group (10%) | Parotid_L (75%) | $D_{mean} \leq 24$ Gy |
| | Parotid_R (25%) | $D_{mean} \leq 37$ Gy |

*In the audit methodology the defined tolerance dose is $D_{98\%} \geq 90\%$. To make it stricter, we have opted by considering $D_{98\%} \geq 95\%$, in this study.

As indicated by the equations (4.6) and (4.7), the structure score is equal to 1 if the dosimetric goal (tolerance dose) was achieved, less than 1 if the plan performed better than the tolerance dose for that structure, and higher than 1 if the dosimetric goal was not met. Thus, the lower the global plan score, the better the plan quality [23].

Besides the score definition, SPIDERplan also displays all this information by means of customized radar plots. The circular area is divided into sectors (3 in this case), being their angular

amplitude proportional to the defined group weights. Subsectors within each of these 3 sectors reflect the structures relative weights inside the group. Each structure score is marked as a point within the corresponding subsector at a distance from the centre equal to its value and the polygon resulting from the connection of all points represents the overall plan quality. If a plan does not meet the proposed dose-volume constraints for a given structure, the corresponding vertex of the polygon comes outside the inner circle that has a radius equal to 1. The closer to the centre the polygon is, the higher its plan quality and the lower its global plan score [23].

### 4.2.5 | Correlation between plan complexity, quality and audit measurement results

The correlation between the normalized plan complexity score, plan quality score and the audit measurement results was investigated using the Spearman's correlation coefficients, $r_s$. The strength of the correlations was evaluated as described in Section 4.2.2, with a significance level of 5%. In terms of deliverability, the correlation between the local pre-treatment QA verification results and the on-site audit measurements was also assessed.

## 4.3 | Results

### 4.3.1 | Plan deliverability

Results of the dosimetric verification of the IMRT/VMAT audit plans, previously published [3], showed that all audit plans met the established acceptance criteria. Ionization chamber measurements were within the audit tolerances of ±5% for PTVs and ±7% for the spinal cord in all institutions, being below ±3% in most of them. For film analysis, gamma passing rates were on average 96.9±3.0% ranging from 90.3% to 99.1%, all above the acceptance limit of 90% for a criterion of 3% global dose/3 mm, 20% TH. Only four institutions had suboptimal results, i.e., gamma passing rates between 90% and 95% and IC differences above 2% in PTVs. Those institutions used different IMRT techniques (SW, SS, and VMAT) and TPSs (Monaco, XiO, Eclipse) [3]. When a more stringent criterion (2%/2 mm 20% TH) was utilized for gamma analysis, a wider range of passing rates was obtained (mean±SD: 86.6±8.9%), with 9 institutions having results below 90%.

The audit plans deliverability was also verified using the institution's pre-treatment verification equipment, evaluation metrics and acceptability criteria. Gamma analysis was the most widely employed comparison method and 95% (3%/3 mm) the acceptability criteria adopted in the majority of the institutions. The obtained gamma passing rates were on average 98.9±1.1%. No correlation between the local patient-specific QA and audit measurement results was identified.

### 4.3.2 | Plan complexity metrics

The audit plans presented different characteristics and levels of complexity, despite having all met the prescription conditions. For instance, even if most institutions used VMAT as the treatment technique, the number of arcs varied between 2 (n=10) and 4 (n=3), with two institutions using 3 arcs. All SW plans used 9 beams, while the SS had 7. A great difference (about 2 times, on average) between the total MU for SW plans in comparison to VMAT ones as well as a large spread in total MUs of VMAT Eclipse plans were noticed. For VMAT Eclipse plans, the average number of MUs was 667.8±164.3, and 775.9±73.4 for VMAT Monaco. The SW Eclipse plans had 1548.0±49.5 while the SW Monaco had 1215 MU. The SS plan had 843.6 MU.

Concerning complexity metrics, Figure 4.1 represents four pairs (PM vs 1-MCS, SAS20 vs SAS10, PI vs EM, and MAD vs CAS) of computed indices. As can be seen, some metrics seem correlated and allow to separate different groups of technology. SW Eclipse plans are the ones with the highest modulation (PM, 1-MCS) – Figure 4.1a) – and the smallest openings (SAS10, SAS20) – Figure 4.1b), while VMAT Eclipse plans have the most irregularly shaped apertures (PI and EM) – Figure 4.1c). Regarding the asymmetry, when looking at the mean asymmetry distance (MAD), Elekta TPS (Monaco and XiO) plans and non-VMAT plans, tend to have the highest values – Figure 4.1d). Generally, a wide variability in VMAT Eclipse plans can be observed.

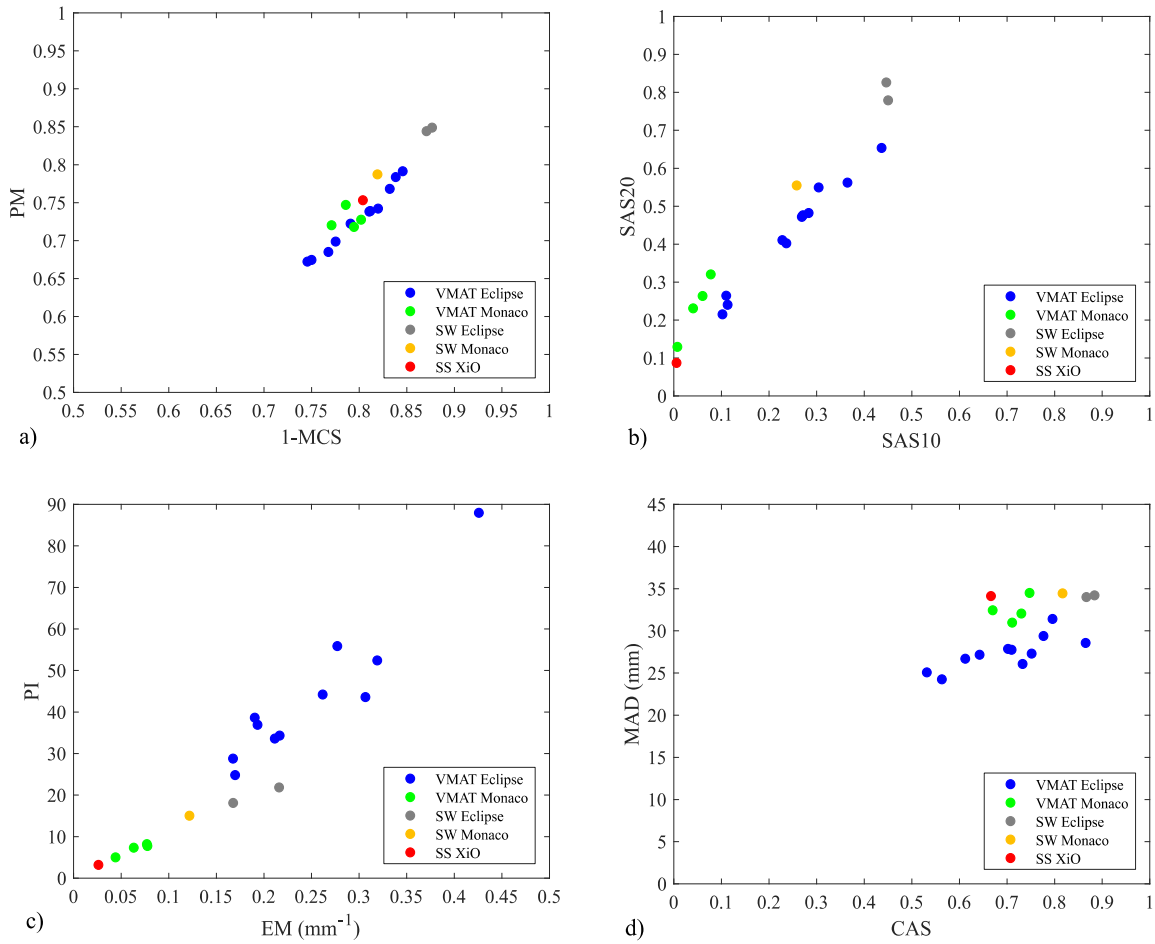


**Figure 4.1 –** Representation of four pairs of complexity metrics. Each data point corresponds to an H&N audit plan.

Figure 4.1 points to the existence of correlation between some of the computed metrics. To further explore it, the Spearman's correlation coefficients $r_s$ and corresponding p-values p were determined and are presented in Figure 4.2. As the correlation matrix shows, metrics that assess plan modulation, namely MU, 1-MCS and PM are strongly correlated ($r_s$>0.6, p<0.006). Regarding SAS10 and SAS20 that give information about the percentage of small leaf openings, a very strong dependence ($r_s$=0.951, p=0.000) was found, as well as between PI and EM, that quantify the beam apertures irregularity ($r_s$=0.937, p=0.000). MAD and CAS that are related with the position of the beam apertures, present a moderate dependency ($r_s$=0.586, p=0.010).

Concerning the correlations between each complexity metric and plan deliverability, Table 4.2 presents the obtained results. The total plan MUs present a moderate correlation with the average IC % difference, though not statistically significant ($r_s$=0.465, p=0.047). We can thus say that plans

with a larger number of MU tended to be associated with higher deviations between calculated and measured doses. The use of a more stringent criteria in gamma analysis did not contribute to any evident difference in the correlation analysis.

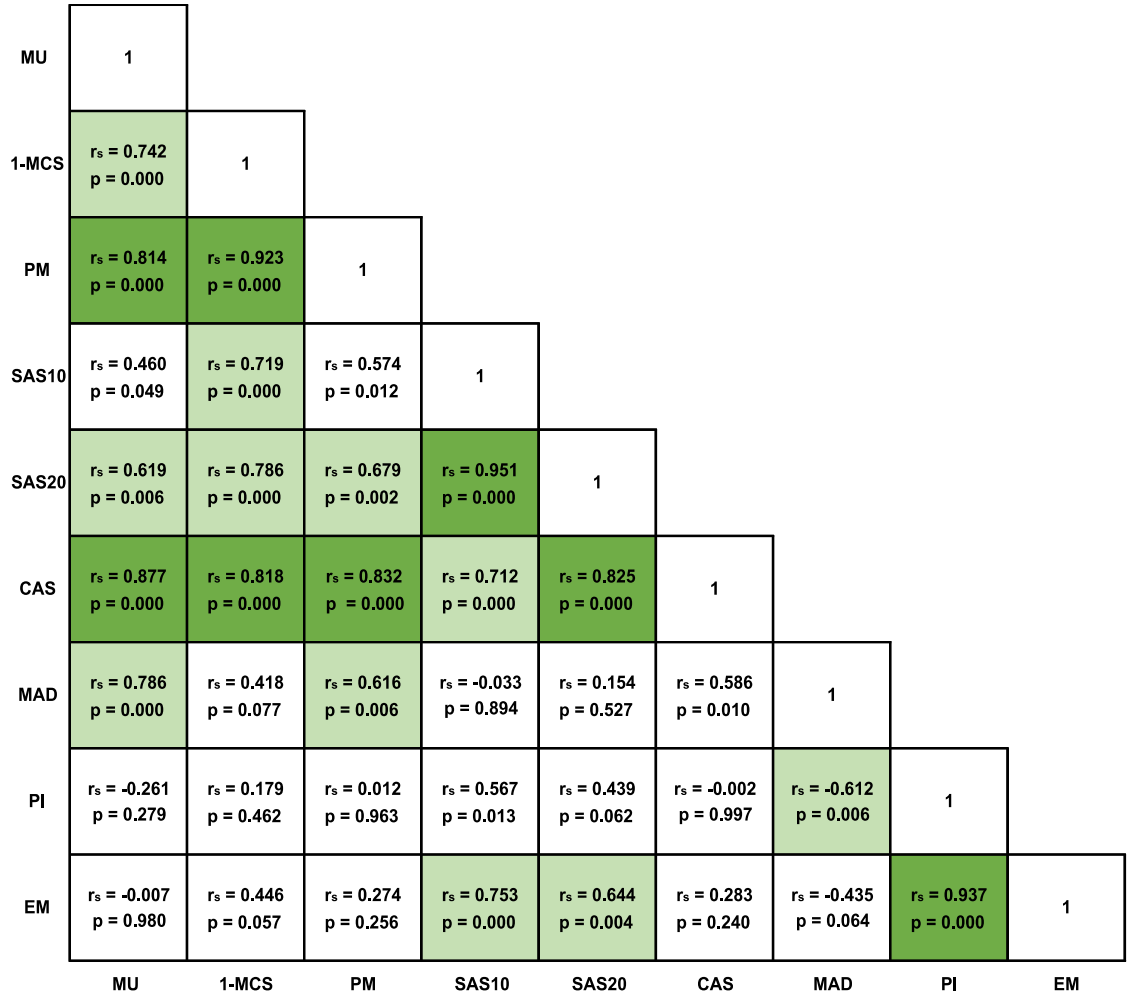


**Figure 4.2 –** Spearman's correlation coefficients $r_s$ and corresponding p-values. Strong and very strong correlations are marked in light and dark green, respectively. Correlations are considered significant for $p < 0.006$ (Bonferroni-corrected).

**Table 4.2 –** Spearman's rank correlation coefficients $r_s$ and their corresponding p-values (within parenthesis) between each complexity metric and the deliverability results. Correlations are considered significant for $p < 0.006$ (Bonferroni-corrected).

| | Average IC % difference | Gamma passing rate 3%/3 mm 20%TH | Gamma passing rate 2%/2 mm 20%TH |
|---|---|---|---|
| **MU** | 0.465 (0.047) | -0.201 (0.409) | -0.211 (0.385) |
| **1-MCS** | 0.274 (0.256) | 0.055 (0.824) | 0.125 (0.610) |
| **PM** | 0.384 (0.105) | 0.044 (0.858) | 0.133 (0.585) |
| **SAS10** | 0.040 (0.871) | 0.157 (0.521) | 0.191 (0.431) |
| **SAS20** | 0.212 (0.381) | 0.000 (1.000) | 0.044 (0.860) |
| **CAS** | 0.268 (0.265) | 0.029 (0.906) | 0.088 (0.721) |
| **MAD** | 0.367 (0.123) | -0.090 (0.714) | -0.056 (0.820) |
| **PI** | 0.044 (0.860) | 0.007 (0.977) | 0.144 (0.555) |
| **EM** | 0.081 (0.743) | 0.044 (0.858) | 0.218 (0.369) |

### 4.3.3 | Plan complexity score

From the computed complexity metrics, PCA was performed to obtain a new set of uncorrelated variables (principal components, PCs), being the corresponding loadings and cumulative percentage of the total explained variance presented in Table 4.3.

**Table 4.3 –** Loadings and cumulative % of the total explained variance corresponding to the calculated PCs.

| | **PC1** | **PC2** | **PC3** | **PC4** | **PC5** | **PC6** | **PC7** | **PC8** | **PC9** |
|---|---|---|---|---|---|---|---|---|---|
| **Total MU** | **0.371** | -0.261 | -0.187 | -0.056 | 0.592 | -0.117 | 0.594 | -0.096 | -0.173 |
| **1-MCS** | **0.405** | -0.035 | 0.248 | -0.494 | -0.431 | -0.489 | 0.085 | -0.289 | 0.117 |
| **PM** | **0.400** | -0.134 | 0.266 | -0.426 | 0.003 | 0.641 | -0.115 | 0.366 | -0.110 |
| **SAS10** | **0.371** | 0.257 | -0.362 | 0.040 | 0.048 | -0.424 | -0.327 | 0.575 | -0.209 |
| **SAS20** | **0.395** | 0.152 | -0.411 | 0.025 | 0.166 | 0.255 | -0.421 | -0.540 | 0.300 |
| **CAS** | **0.395** | -0.113 | -0.090 | 0.595 | -0.558 | 0.199 | 0.253 | -0.048 | -0.227 |
| **MAD** | 0.212 | **-0.467** | 0.508 | 0.413 | 0.264 | -0.222 | -0.374 | 0.044 | 0.217 |
| **PI** | 0.096 | **0.555** | 0.464 | 0.114 | 0.216 | -0.003 | -0.086 | -0.298 | -0.558 |
| **EM** | 0.186 | **0.533** | 0.227 | 0.176 | 0.080 | 0.054 | 0.361 | 0.241 | 0.633 |
| **cumulative % of total variance** | **60.594** | **90.983** | 95.901 | 98.037 | 99.410 | 99.656 | 99.806 | 99.945 | 100 |

The first two PCs together explain over 90% of the total variation in the dataset, 60.594% and 30.389%, respectively, which is above the predefined cut-off (80%). All loadings in PC1 are positive, meaning that the PC increases with the increasing values of all complexity metrics. Therefore, this PC can be interpreted as an indicator of the plan global modulation. The variables that contribute the most for PC1 are: the total MU, 1-MCS, PM, SAS10, SAS20, and CAS, as they have the highest loadings. In PC2, the loadings have both positive and negative signs, which was expected because PC1 has only positive signs – orthogonality condition explained in Section 4.2.3. PI, EM and MAD are the variables with the highest loadings, but while the value of PC2 increases with increasing PI and EM, decreases with MAD. Therefore, PC2 shows a contrast between PI and EM that evaluate the apertures irregularity, and MAD that evaluate their asymmetry.

Figure 4.3 presents a scatter-plot of PC2 vs PC1 for the treatment complexity data, where each data point represents an H&N audit plan (primary axis). The PC loading values for each variable are overlaid on the same figure, represented by a vector (secondary axis). The x-component of the vector is the loading of that variable in PC1 and the y-component is the loading associated with PC2. This kind of representation is named biplot.

In Figure 4.3, plans represented by a point near the origin have values close to the mean for all variables. For PC1, less complex plans are represented on the left side and the more complex on the right part of the graph. Regarding PC2, less complex plans are shown towards the bottom and the more complex towards the top.

As mentioned above, all loadings of PC1 are positive, and that is clearly visible, as all vectors point to the right side of the graph. The difference in the sign between the PI, EM (towards the top right of the plot) and MAD (towards the bottom right of the plot) in PC2 is also evident. From Figure 4.3 analysis, it becomes also clear that MU, PM, CAS, 1-MCS, SAS10 and SAS20 are the variables that contribute the most for PC1, given the small angle that the corresponding vectors make with the horizontal axis, and that PI, EM and MAD are the variables that contribute the most for PC2. The correlation between any two variables can be evaluated by the corresponding vectors angle. If the correlation between two variables is positive and strong, the angle between the vectors is small (e.g. PM vs CAS), if the correlation is weak, the angle is around 90° (eg. MU vs EM), and if the correlation is negative and strong (eg. MAD vs PI) the angle is closer to 180°.

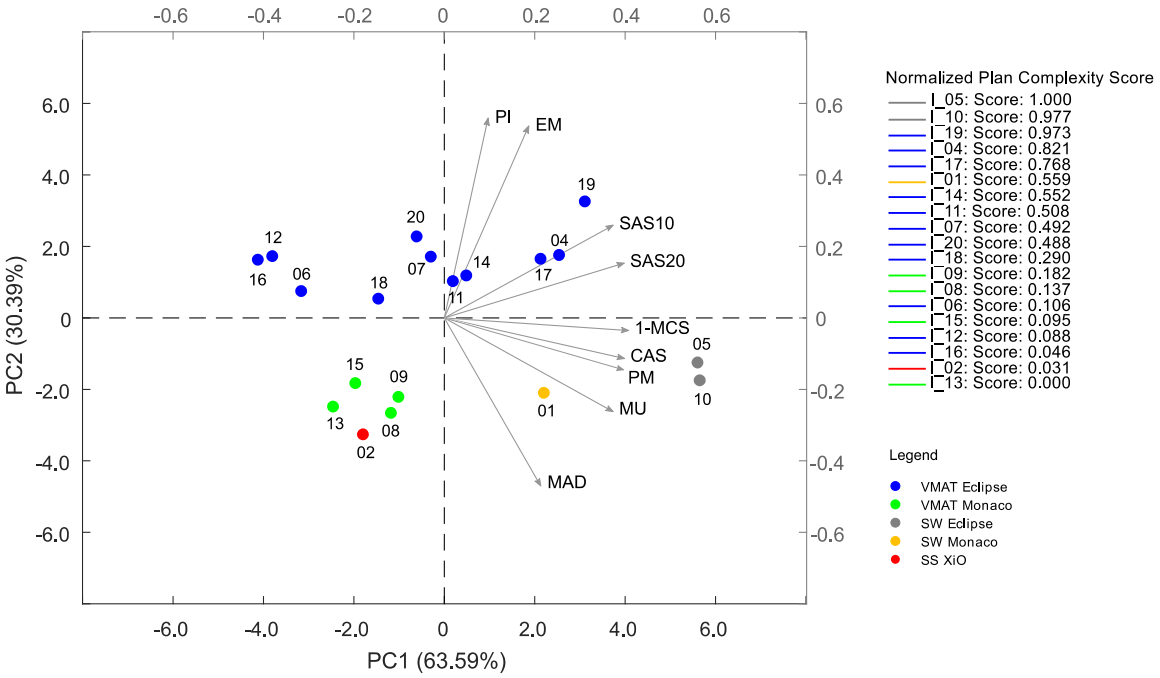


**Figure 4.3** – Biplot for the treatment plans complexity data. Each data point represents an H&N audit plan (primary axis), being the corresponding ID displayed next to it. The loadings of each variables are represented by vectors (secondary axis). The normalized plan complexity score (nPCS) is presented on the right, for each of the 19 audited institutions as labelled by I_XX, where XX is the number of the institution.

Looking now at the distribution of the points representing each plan, different groups can be clearly identified. Group I, on the left lower quadrant, includes VMAT Monaco plans, and the SS XiO plan, having a negative PC1 and PC2. Group II, on the right lower quadrant, comprises the SW plans, which have positive PC1 and negative PC2. Group III, in the top, encompasses all VMAT Eclipse plans and have a positive PC2. This group can be further divided into two groups, one with positive PC1, and another with negative PC1. The higher the values of each PC, the greater the complexity, which means that SW Eclipse plans are the most complex for PC1, and VMAT Eclipse plans the most complex for PC2. In other words, SW Eclipse plans have higher global modulation, evaluated by PC1, while VMAT Eclipse plans have more irregularly shaped apertures (PI and EM) and lower mean asymmetric distances (MAD), evaluated by PC2.

To summarize and give a global meaning to the information provided in the biplot, the plan complexity score was calculated as the weighted of the first two PCs. The normalized version of the PCS (nPCS) was used to rank of plans according to their relative complexity, being also presented in Figure 4.3 (on the right side). According to the definition, the plan with minimum PCS is from institution #13, a VMAT Monaco plan, and the most complex is from institution #5, an SW Eclipse plan. The two SW plans created in Eclipse are indeed the most complex audit plans. VMAT Eclipse plans are amongst the most complex plans (top), but also amongst the less complex ones (bottom), which illustrates the variability that can be observed for plans created using the same technique and technology.

### 4.3.4 | Plan quality

The plan diagram obtained for the plans using SPIDERplan is presented in Figure 4.4, as well as the quality global plan score. In the radar plot, each polygon represents a different plan and the color code is the same as in the previous figures.

From Figure 4.4 analysis, it is visible that all institutions had a similar quality performance in the PTVs, being the spread higher for the organs-at-risk, mainly in the spinal cord and right parotid. The global plan score, on the right of the image, indicates that apart from the two best plans, the treatment plans quality did not differ much among institutions, with a great number of institutions (n=10) having a global plan score between 0.921 and 0.940. The VMAT plans were generally better, with the top 5 occupied by 4 VMAT Eclipse and 1 VMAT Monaco plan.

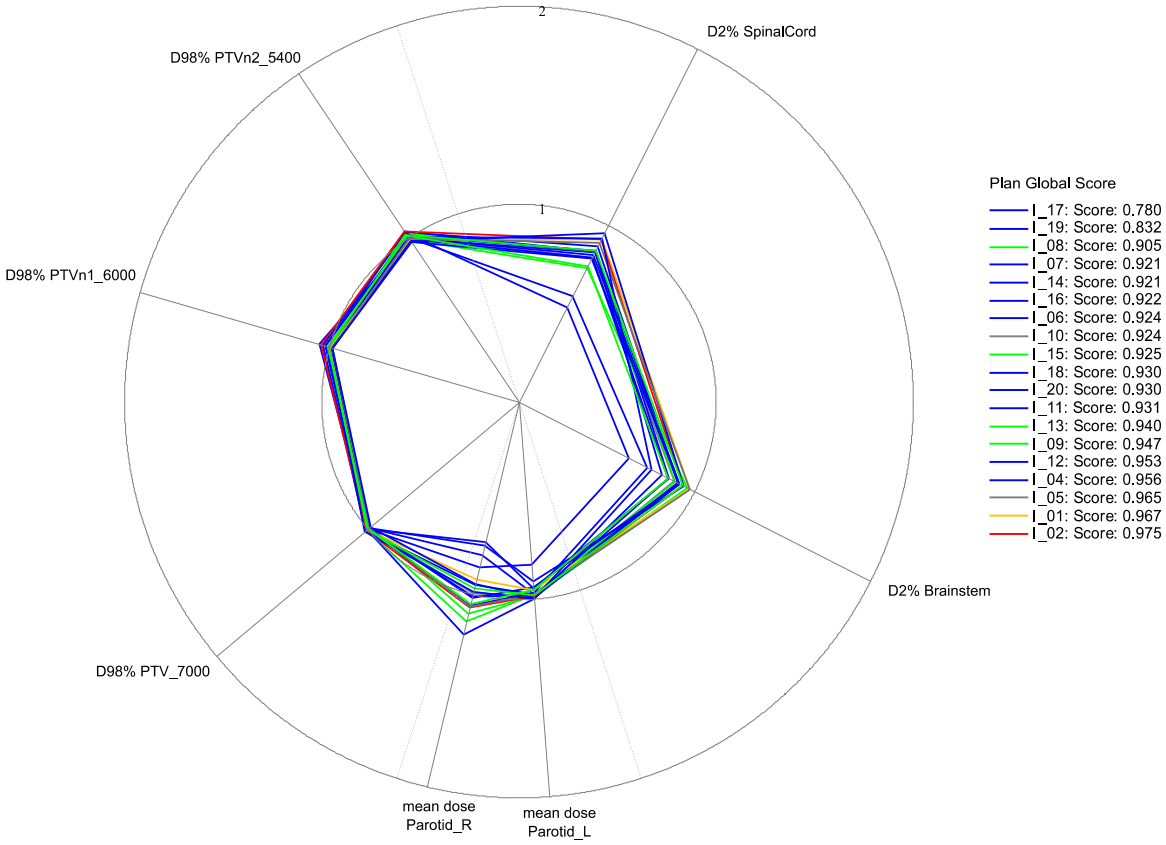


**Figure 4.4 –** Plan diagram obtained in SPIDERplan. Each polygon represents the quality of a different plan. The global plan score is presented on the right side, for the 19 audited institutions. The color code is the same as in previous figures: VMAT Eclipse (blue), VMAT Monaco (green), SW Eclipse (gray), SW Monaco (yellow), SS XiO (red).

### 4.3.5 | Correlation between the plan complexity, plan quality and audit measurements

The correlations between the plan complexity score, the plan quality and deliverability were investigated through the calculation of the Spearman's rank correlation coefficients. Starting with the analysis of the dependency between plan complexity and quality, Figure 4.5 presents the global plan score versus the normalized plan complexity score for the audit plans. The Spearman's correlation coefficient and corresponding p-values are also shown.

No significant correlation between plan quality and complexity has been identified. As can be seen in Figure 4.5 and as mentioned above, the quality of the audit plans did not differ much as opposed to the plans complexity, which explains the lack of correlation. Yet, it is of note that the two plans having a remarkably better quality (lower global plan score) are among the most complex plans.

The relationship between plan complexity and deliverability, considering each PC (PC1 and PC2) separately and the global score, was also investigated, but no significant correlations were identified. Regarding the analysis of the relationship between plan quality and deliverability, the global plan score and the film gamma passing rates presented a significant correlation, for both 3%/3 mm 20% TH, $r_s$=-0.488, p=0.034, and 2%/2 mm 20% TH criteria, $r_s$=-0.560, p=0.013. This means that plans with a higher quality tended to perform better in terms of gamma passing rates.

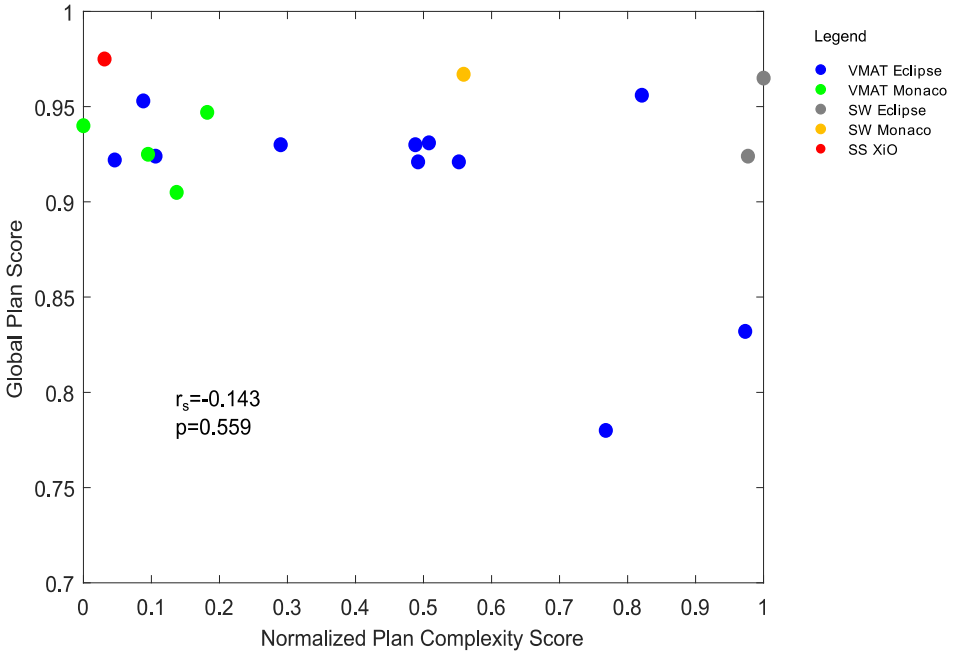


**Figure 4.5 –** Global plan score versus the normalized plan complexity score for the audit plans.

## 4.4 | Discussion

The calculation of multiple metrics makes it possible to assess different complexity aspects of treatment plans. However, having an overview of the global complexity and the differences between groups of technologies/techniques in a simple way is not usually feasible, especially when considering a large number of metrics. In this study, multiple complexity metrics have been computed for the IMRT/VMAT plans of a national audit project. An approach based on principal component analysis was followed to reduce the dimensionality of the original dataset to their most descriptive components, allowing the visualization of the information in two dimensions. Moreover, a global plan complexity score was proposed to summarize all complexity aspects associated with a plan, based on a methodology already followed to calculate socio-economic status indices [26,27]. The main advantage of computing a global score over the commonly used approach of analysing each metric individually (or in pairs) is that only independent variables are considered in its calculation, being much easier to find common trends. Moreover, this score allowed to rank the plans according to their complexity, which can be a useful application in benchmarking exercises, such as multi-institutional studies/audits.

PCA has already been utilized in the context of plan complexity analysis, playing different roles. Chun et al [28] investigated the impact of some modulation indices and plan parameters on the results of pre-treatment verification considering a set of 758 VMAT plans. In that study, PCA was used to examine the dependencies between variables. Valdes et al [29] developed an algorithm to predict a priori the results of the patient-specific QA verification. This study involved 498 sliding

window plans and their complexity was characterized by 78 metrics. Due to the large amount of data, PCA was used to facilitate data visualization.

In the present study, a set of 19 IMRT/VMAT plans created in the context of a national audit and 9 complexity metrics were considered for PCA analysis. The reduced number of plans does not represent a limitation for PCA application, despite larger sample sizes tend to be considered more robust [30–32]. The first two PCs explained over 90% of the variability in the dataset. Therefore, the representation of PC2 vs PC1 is a good approximation of the original scatter plot in 9 dimensions, and allowed to identify groups that were clearly separated. SW Eclipse plans had higher global modulation, evaluated by PC1, while VMAT Eclipse plans had the most irregularly shaped apertures and lowest mean asymmetry aperture distances, evaluated by PC2.

The global plan complexity score computed as the weighted mean of the PC1 and PC2, summarized the information provided in the two-dimensional representation of the PCs, allowing to compare the relative complexity of the audit treatment plans and rank the plans accordingly – Figure 4.3. In the top 5 there were two SW Eclipse and three VMAT Eclipse plans. VMAT Eclipse plans were amongst the most complex plans (top), but also amongst the less complex ones (bottom), which illustrates the variability that can be observed between plans created using the same technique and technology. This variability can be attributed to differences in planning strategies and/or planner's skills. The VMAT Monaco plans are generally less complex than the VMAT Eclipse ones. As the computed metrics were all aperture-related, it can be concluded that VMAT Eclipse tend to put more demand on the MLC when compared with VMAT Monaco plans. These findings are consistent with the ones reported by Hernandez et al [15]. According to the authors, VMAT Monaco plans can reduce the MLC demand as they further modulate the gantry speed and the dose rate. Even though no evaluation of the gantry speed and dose rate variations was included in the present study, this was confirmed.

No clear differences between the quality of the plans created using different technologies and treatment techniques were identified in this study. Indeed, apart from plans from institutions #17 and #19, the plans quality did not differ much. Both institutions #17 and #19, used Eclipse TPS and VMAT. Institution #19 was one of the 3 institutions that created a VMAT plan with 4 arcs and this was by far the most complex VMAT plan, as indicated by the normalized PCS. Arcs 3 and 4 were particularly complex, with much smaller and irregular segments given a high weight, in comparison with other VMAT plans. Institution #17 plan had 2 arcs, and it is also in the top of the most complex. This institution put a great demand on the spinal cord in the optimization phase, as they typically do in clinical practice, and the difference in the near-max dose (D2%) is notorious in Figure 4.4. In spite of the high level of complexity of these two plans, no issues regarding deliverability were identified. However, this was not the case of institution #4. Institution #4 created a plan with 4 arcs that was among the most complex plans. But, the 'excess' of complexity in this case did not contribute to obtain an improved dose distribution. Moreover, this institution was in the group of institutions presenting suboptimal results, with a gamma passing rate of 91.3% and IC average deviation of 3.1%. Together with other factors that were identified in a previous publication [3] and that will be mentioned below, the plan complexity may have contributed to the suboptimal results. The other institution that had a plan with 4 arcs, was institution #9. This was the most complex plan among the VMAT Monaco group but it was also the one in that group with poorer quality and with the highest audit measurement deviations – gamma passing rate of 95.1% and IC average deviation of 2.6%. To summarize, this comparison drives the attention to two important aspects. First, it corroborates the idea that higher complex plans do not necessarily imply improved dose distributions. Second, that an increased complexity may compromise the treatment deliverability, and ultimately the treatment outcome [15,16].

The results of the correlation analysis between complexity and audit measurements showed that only the total number of MU and the average percent ionization chamber deviation presented a moderate dependency. Plans with larger number of MU tended to be associated with a higher difference between calculated and measured doses.

The total plan MU is the simplest and at the same time the most widely used indicator of complexity. In the context of the IMRT/VMAT audit methodology, the total number of MU was used to classify the plans complexity in different levels, depending on the treatment technique as 'High', 'Medium', and 'Low' [24,33]. In this work, a more detailed analysis was performed and the results generally agreed with the assessment based on the total MU. To illustrate this, the VMAT plans were considered. Starting with the largest group, the VMAT Eclipse, plans from institutions #4 and #17 were classified as 'High' complex, based on the total MU, being among the most complex for this group, as indicated by the normalized PCS – Figure 4.3. Complexity of the plans from institutions #6, #12 and #16 was classified as 'Low' and these are indeed the VMAT Eclipse plans with the lowest normalized PCS. Considering the four VMAT plans created in Monaco TPS, one of them was considered 'High' complex – institution #9 – and the other three as 'Medium'. The plan from institution #9 has shown to be the most complex VMAT plan for this group, but it is less complex than most of the VMAT Eclipse plans – Figure 4.3. These results generally support the use of the total number of MU as a simple indicator of plan complexity for the purpose of the audit to alert the local medical physicists' team for possible over-modulation. Nevertheless, the reported findings demonstrate the importance of performing a more comprehensive analysis when comparing plans created using different technologies, even for the same treatment technique.

The lack of significant correlations in multi-institutional studies was also reported by other authors. Hernandez et al [15] in a study involving 40 audit and 100 clinical VMAT plans from various institutions, did not find any correlation between plan quality and complexity, nor between plan complexity and the results of independent QA measurements. Glenn et al [18] could neither identify any dependency between a single complexity metric out of the 16 computed and the results of IROC H&N phantom irradiations. On the other hand, McGarry et al [17] in an analysis of the complexity of 39 VMAT plans resulting from a dosimetry audit in the UK, showed that both MCS and MU of plans created to be delivered by Varian linacs, were correlated with the results of the QA phantom irradiations. MU was also significantly correlated with the plan quality for another specific group of technology.

In dosimetry audits, the clinical cases, planning objectives, verification tools, and evaluation methods tend to be standardized. However, there are many factors that may contribute to increased uncertainties in different ways making it difficult to find strong correlations between audit measurement results and complexity. Generally, some of those factors are related with dose calculation in TPS (beam modelling, commissioning inaccuracies), phantom positioning verification or machine performance [18]. The verification devices and evaluation methods and criteria, mainly gamma analysis, may also not be sensitive enough, as reported by some authors [34–36]. More specifically, in the national IMRT/VMAT audit conducted in Portugal, some factors were identified that may had contributed to increased uncertainty in dose delivery, which could have masked the impact of plan complexity on deliverability. Those factors included: no consideration of the treatment couch (11/19 institutions did not account for the couch in treatment planning), equipment age (6/19 linacs were more than 10 years old) and small beams modelling in the TPS (in 4/19 institutions the calculated output factors for 3×3 $cm^2$ and/or 2×2 $cm^2$ fields were out of the tolerance, when comparing with a reference dataset [37,38]) [3].

In this work, PCA was the starting point to calculate a global plan complexity score. The proposed score is relatively easy to compute but its construction involves some phases of subjective analysis, such as the initial selection of the complexity metrics, the approach followed to deal with

missing values (if any) and the decision of how many PCs to retain. The original variables may also need to be initially standardized as variance is affected by the data scale. This makes the PCA results sensitive to modifications of the original dataset, such as the inclusion of new variables and/or new plans. Moreover, this score is a relative measure of the complexity between the considered plans, not providing information on absolute level of complexity nor on the potential impact of each metric on plan deliverability. The subjacent analysis is useful to understand how the various metrics contribute to the differences between plans, and if any of them is strongly associated with the plan deliverability. The proposed plan complexity score should be interpreted as a single indicator that summarizes all the evaluated aspects, allowing to compare in an easy way any set of plans, summarizing similarities and differences.

The applied statistical approach suggested that it is possible to simplify the plans complexity analysis and comparison among groups, but still further investigation of its utility and applicability using a large set of plans is needed.

## 4.5 | Conclusion

The complexity of the H&N IMRT/VMAT plans created during an on-site audit carried out in Portugal in 2018 was evaluated and compared for the different groups of technologies and IMRT/VMAT treatment techniques involved in the project. A statistical approach based on principal component analysis was followed to simplify the analysis of the complexity of the plans, allowing to clearly identify groups of different techniques and technologies. Based on the PCA results, a global plan complexity score was proposed to characterize the complexity through a single indicator. This global score summarized the differences in complexity between and within the different groups of audit plans, paving the way for improvement of the planning strategies at the national level through knowledge sharing. Moreover, the proposed methodology can be applied to any given plans set and the resulting scores may be taken as indicators to compare the global complexity among those plans.

## 4.6 | Funding

This work was supported by FCT, the Portuguese Foundation for Science and Technology, through a PhD scholarship, reference SFRH/BD/118929/2016.

## 4.7 | 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.

## 4.8 | Acknowledgments

The authors would like to thank Francisco Caramelo, Joana Bastos, Pedro Sá-Couto and Andreia Hall for their assistance with the statistical analysis.

## 4.9 | References


[1] Fenkell L, Kaminsky I, Breen S, Huang S, Van Prooijen M, Ringash J. Dosimetric comparison of IMRT vs. 3D conformal radiotherapy in the treatment of cancer of the cervical esophagus. Radiother Oncol 2008;89:287–91. https://doi.org/10.1016/j.radonc.2008.08.008.

[2] Nutting CM, Morden JP, Harrington KJ, Urbano TG, Bhide SA, Clark C, et al. Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): A phase 3 multicentre randomised controlled trial. Lancet Oncol 2011;12:127–36. https://doi.org/10.1016/S1470-2045(10)70290-4.

[3] Santos T, Lopes MC, Gershkevitsh E, Vinagre F, Faria D, Carita L, et al. IMRT national audit in Portugal. Phys Medica 2019;65:128–36. https://doi.org/10.1016/j.ejmp.2019.08.013.

[4] Mohan R, Arnfield M, Tong S, Wu Q, Siebers J. The impact of fluctuations in intensity patterns on the number of monitor units and the quality and accuracy of intensity modulated radiotherapy. Med Phys 2000;27:1226–37. https://doi.org/10.1118/1.599000.

[5] Crowe SB, Kairn T, Kenny J, Knight RT, Hill B, Langton CM, et al. Treatment plan complexity metrics for predicting IMRT pre-treatment quality assurance results. Australas Phys Eng Sci Med 2014;37:475–82. https://doi.org/10.1007/s13246-014-0274-9.

[6] McNiven AL, Sharpe MB, Purdie TG. A new metric for assessing IMRT modulation complexity and plan deliverability. Med Phys 2010;37:505–15. https://doi.org/10.1118/1.3276775.

[7] Llacer J, Solberg TD, Promberger C. Comparative behaviour of the Dynamically Penalized Likelihood algorithm in inverse radiation therapy planning. Phys Med Biol 2001;46:2637. https://doi.org/10.1088/0031-9155/46/10/309.

[8] Webb S. Use of a quantitative index of beam modulation to characterize dose conformality: Illustration by a comparison of full beamlet IMRT, few-segment IMRT (fsIMRT) and conformal unmodulated radiotherapy. Phys Med Biol 2003;48:2051–62. https://doi.org/10.1088/0031-9155/48/14/301

[9] Du W, Cho SH, Zhang X, Hoffman KE, Kudchadker RJ. Quantification of beam complexity in intensity-modulated radiation therapy treatment plans. Med Phys 2014;41. https://doi.org/10.1118/1.4861821.

[10] Younge KC, Matuszak MM, Moran JM, McShan DL. Penalization of aperture complexity in inversely planned volumetric modulated arc therapy. Med Phys 2012;39:7160–70. https://doi.org/10.1118/1.4762566.

[11] Park JM, Park SY, Kim H, Ho Kim J, Carlson J, Ye SJ. Modulation indices for volumetric modulated arc therapy. Phys Med Biol 2014;59:7315–40. https://doi.org/10.1088/0031-9155/59/23/7315.

[12] Masi L, Doro R, Favuzza V, Cipressi S, Livi L. Impact of plan parameters on the dosimetric accuracy of volumetric modulated arc therapy. Med Phys 2013;40. https://doi.org/10.1118/1.4810969.

[13] Chiavassa S, Bessieres I, Edouard M, Mathot M, Moignier A. Complexity metrics for IMRT and VMAT plans: a review of current literature and applications. Br J Radiol 2019;92:20190270. https://doi.org/10.1259/bjr.20190270.

[14] Antoine M, Ralite F, Soustiel C, Marsac T, Sargos P, Cugny A, et al. Use of metrics to quantify IMRT and VMAT treatment plan complexity: A systematic review and perspectives. Phys Medica 2019;64:98–108. https://doi.org/10.1016/j.ejmp.2019.05.024.

[15] Hernandez V, Saez J, Pasler M, Jurado-Bruggeman D, Jornet N. Comparison of complexity metrics for multi-institutional evaluations of treatment plans in radiotherapy. Phys Imaging Radiat Oncol 2018;5:37–43. https://doi.org/10.1016/j.phro.2018.02.002.

[16] Jurado-Bruggeman D, Hernandez V, Saez J, Navarro D, Pino F, Martinez T, et al. Multi-centre audit of VMAT planning and pre-treatment verification. Radiother Oncol 2017;124:302–10. https://doi.org/10.1016/j.radonc.2017.05.019.

[17] McGarry CK, Agnew CE, Hussein M, Tsang Y, McWilliam A, Hounsell AR, et al. The role of complexity metrics in a multi-institutional dosimetry audit of VMAT. Br J Radiol 2016;89. https://doi.org/10.1259/bjr.20150445.

[18] Glenn MC, Hernandez V, Saez J, Followill DS, Howell RM, Pollard-Larkin JM, et al. Treatment plan complexity does not predict IROC Houston anthropomorphic head and neck phantom performance. Phys Med Biol 2018;63. https://doi.org/10.1088/1361-6560/aae29e.

[19] Gurney-Champion OJ, Collins DJ, Wetscherek A, Rata M, Klaassen R, van Laarhoven HWM, et al. Principal component analysis for fast and model-free denoising of multi b-value diffusion-weighted MR images. Phys Med Biol 2019;64:105015. https://doi.org/10.1088/1361-6560/ab1786.

[20] Benestad R, Parding K, Dobler A, Mezghani A. A strategy to effectively make use of large volumes of climate data for climate change adaptation. Clim Serv 2017;6:48–54. https://doi.org/10.1016/j.cliser.2017.06.013.

[21] Tsiamas P, Bagher-Ebadian H, Siddiqui F, Liu C, Hvid CA, Kim JP, et al. Principal component analysis modeling of Head-and-Neck anatomy using daily Cone Beam-CT images. Med Phys 2018;45:5366–75. https://doi.org/10.1002/mp.13233.

[22] Jolliffe IT, Cadima J, Cadima J. Principal component analysis: a review and recent developments Subject Areas. PhilTransRSocA 2016;374:1–16. https://doi.org/10.1098/rsta.2015.0202.

[23] Ventura T, Lopes MC, Ferreira BC, Khouri L. SPIDERplan: A tool to support decision-making in radiation therapy treatment plan assessment. Reports Pract Oncol Radiother 2016;21:508–16. https://doi.org/10.1016/j.rpor.2016.07.002.

[24] Kazantsev P, Lechner W, Gershkevitsh E, Clark CH, Venencia D, Van Dyk J, et al. IAEA methodology for on-site end-to-end IMRT/VMAT audits: an international pilot study. Acta Oncol 2019:1–8. https://doi.org/10.1080/0284186X.2019.1685128.

[25] Lee N, Garden A, Kim J, Mechalakos J, Pfister DG, Ang KA, et al. A phase II study of concurrent chemoradiotherapy using three-dimensional conformal radiotheray (3D-CRT) or Intensity-Modulated Radiation Therapy (IMRT) + Bezacizumab (BV) for locally or regionally advanced nasopharyngeal cancer. NRG Oncology – RTOG 0615 2014.

[26] Vyas S, Kumaranayake L. Constructing socio-economic status indices: how to use principal components analysis. Heath Policy Plan 2006; 21:459-68. https://doi.org/10.1093/heapol/czl029.

[27] Antony GM, Rao KV. A composite index to explain variations in poverty, health, nutritional status and standard of living: Use of multivariate statistical methods. Public Health 2007:578–87. https://doi.org/10.1016/j.puhe.2006.10.018.

[28] Chun M, Joon An H, Kwon O, Oh DH, Park JM, Kim J-I. Impact of plan parameters and modulation indices on patient-specific QA results for standard and stereotactic VMAT. Phys Med 2019;62:83-94. https://doi.org/10.1016/j.ejmp.2019.05.005.

[29] Valdes G, Scheuermann R, Hung CY, Olszanski A, Bellerive M, Solberg TD. A mathematical framework for virtual IMRT QA using machine learning. Med Phys 2016;43:4323. https://doi.org/10.1118/1.4953835.

[30] Shaukat S, Rao T, Khan M. Impact of sample size on principal component analysis ordination of an environmental data set: Effects on eigenstructure. Ekológia 2016;35. https://doi.org/10.1515/eko-2016-0014.

[31] Dochtermann N, Jenkins S. Multivariate Methods and Small Sample Sizes. Ethology 2011;117:95–101. https://doi.org/10.1111/j.1439-0310.2010.01846.x.

[32] Jobson JD. Applied Multivariate Data Analysis. vol. II. USA: Springer-Verlag; 1992

[33] International Atomic Energy Agency (IAEA). On-site “end-to-end” dosimetry audits. https://dosimetry-audit-networks.iaea.org/Home/EndToEndAudits [accessed October 25, 2019].

[34] Nelms BE, Zhen H, Tome WA. Per-beam, planar IMRT QA passing rates do not predict clinically relevant patient dose errors. Med Phys 2011;38:1037–44. https://doi.org/10.1118/1.3544657.

[35] Kruse JJ. On the insensitivity of single field planar dosimetry to IMRT inaccuracies. Med Phys 2010;37:2516–24. https://doi.org/10.1118/1.3425781.

[36] Gordon J, Siebers J. Addressing a gap in current IMRT quality assurance. Int J Radiat Oncol Biol Phys 2013;87:20–1. https://doi.org/10.1016/j.ijrobp.2013.03.030.

[37] Followill DS, Kry SF, Qin L, Leif J, Molineu A, Alvarez P, et al. The Radiological Physics Center’s standard dataset for small field size output factors. J Appl Clin Med Phys 2012;13:282–9. https://doi.org/10.1120/jacmp.v13i5.3962.

[38] Followill DS, Kry SF, Qin L, Leif J, Molineu A, Alvarez P, et al. Erratum : "The Radiological Physics Center’s standard dataset for small field size output factors". J Appl Clin Med Phys 2014;15:356–7. https://doi.org/10.1120/jacmp.v15i2.4757.

# CHAPTER 5

# On the complexity of helical tomotherapy treatment plans

Journal of Applied Clinical Medical Physics, 2020, volume 21, issue 7, pages 107-118

T Santos[a,b], T Ventura[b], J Mateus[b], M Capela[b], MC Lopes[b]

[a] Physics Department, University of Coimbra, Coimbra, Portugal
[b] Medical Physics Department, IPOCFG, E.P.E., Coimbra, Portugal

## Abstract

**Purpose:** Multiple metrics are proposed to characterize and compare the complexity of helical tomotherapy (HT) plans created for different treatment sites.

**Methods:** A cohort composed of 208 HT plans from head and neck (105), prostate (51) and brain (52) tumour sites was considered. For each plan, 14 complexity metrics were calculated. Those metrics evaluate the percentage of leaves with small opening times or approaching the projection duration, the percentage of closed leaves, the amount of tongue-and-groove effect and the overall modulation of the planned sinogram. To enable data visualization, an approach based on principal component analysis was followed to reduce the dataset dimensionality. This allowed the calculation of a global plan complexity score. The correlation between plan complexity and pre-treatment verification results using the Spearman's rank correlation coefficients was investigated.

**Results:** According to the global score, the most complex plans were the head and neck tumour cases, followed by the prostate and brain lesions irradiated with stereotactic technique. For almost all individual metrics, head and neck plans confirmed to be the plans with the highest complexity. Nevertheless, prostate cases had the highest percentage of leaves with an opening time approaching the projection duration, whereas the stereotactic brain plans had the highest percentage of closed leaves per projection. Significant correlations between some of the metrics and the pre-treatment verification results were identified for the stereotactic brain group.

**Conclusions:** The proposed metrics and the global score demonstrated to be useful to characterize and quantify the complexity of HT plans of different treatment sites. The reported differences inter- and intra- group may be valuable to guide the planning process aiming at reducing uncertainties and harmonize planning strategies.

**Keywords:** Helical tomotherapy, plan complexity, principal component analysis, plan complexity score

## 5.1 | Introduction

Intensity modulated radiation therapy (IMRT) is becoming a standard treatment technique for cancer patients, using either conventional linear accelerators (linac) or dedicated technologies such as helical tomotherapy (HT). In HT, the radiation beam produced by a compact linac is collimated to a fan-beam. Delivery is done while the beam rotates around the patient and the couch translates through the gantry ring. Both the rotation and translation speeds are constant throughout the treatment. Modulation of beam intensity is accomplished by a pneumatically driven binary multileaf collimator (MLC). An arc-shaped detector array, mounted opposite to the linac, records the exit radiation signal, which can be used for patient positioning verification, plan deliverability evaluation or machine quality assurance (QA) [1].

For treatment planning, parameters like the field width, the pitch and the initial modulation factor are manually set, while each MLC leaf open time per projection (51 by gantry rotation) is determined during the optimization phase. As many authors have reported [2–5], a suboptimal choice of these parameters can compromise plan quality and deliverability, as well as increase treatment time. Therefore, several optimal values and planning approaches have been suggested. For instance, Kissick et al. [4] proposed a rule to choose the pitch values and minimize the longitudinal ripple effect – thread effect – characteristic of HT plans. Shimizu et al. [6] presented a method to derive an initial modulation factor and a site-specific upper limit for this parameter to reduce the delivery time without compromising plan quality. Westerly et al. [2], using a subset of plans with unexpectedly poor pre-treatment QA results, found that these plans had a high percentage of small leaf open times (LOT), being the mean LOT less than 100 ms. After replanning, the mean LOT became higher than 100 ms and the deviations between calculated and measured dose fell within ±3%. This could have happened due to the inaccuracies associated with the modelling of the MLC leaf latency in the treatment planning system (TPS) whose impact is higher for short leaf open times. MLC leaf latency and tongue-and-groove/penumbra effects have indeed been pointed as factors that can affect plan deliverability [2,7].

More comprehensive studies for different treatment sites, including a wider set of TPS reported parameters, such as the couch travel, couch speed, number of gantry rotations, gantry period and treatment time, have been carried out [8,9]. Bresciani et al. [8], using 384 HT plans of multiple treatment sites, found no strong correlations between some of these factors and the results of pre-treatment QA verification. Binny et al. [9] have used multiple statistical process control methods on a set of head and neck (28), pelvic (19) and brain (23) plans, to define lower and upper limits for planning parameters, like the modulation factor, gantry period, and couch speed, based on acceptable pre-treatment QA results. The established ranges were specific to each treatment site and contributed to improve the treatment efficiency at their institution.

Given the numerous degrees of freedom existing in HT, plans created for the same site may have different degrees of complexity, which may not be fully characterized by the TPS reported parameters. The evaluation of radiotherapy plans complexity has been widely researched. Multiple metrics have been proposed for static-gantry IMRT and volumetric modulated arc therapy [10–15]. Complexity analysis has demonstrated to play a role in treatment plans characterization and comparison, contributing to adapt and improve the planning, optimization and QA processes. To date, a comprehensive evaluation of the helical tomotherapy plans complexity, through the definition and extension of some existing metrics is lacking in the literature. Thus, this study aims to quantify, evaluate and compare the complexity of HT plans created for various treatment sites by calculating several metrics. These metrics include some commonly evaluated parameters and novel indices that assess different aspects of the HT plans which may directly or indirectly

contribute to increased uncertainties in dose calculation and delivery. The potential effect of complexity on the plans deliverability was also investigated.

## 5.2 | Materials and Methods

### 5.2.1 | Treatment plans and deliverability evaluation

A total of 208 plans from patients that underwent helical IMRT treatments at our institution were retrospectively analysed. The considered treatment sites included head and neck (105), prostate (51) and brain tumour cases (52). The head and neck plans were generated with simultaneously integrated boost, for two or three dose levels. The prescription dose per fraction to the high-risk planning target volume (PTV) was 2 or 2.12 Gy. In prostate tumour cases, only plans aiming to irradiate the prostate and seminal vesicles or the involved fossa, with a dose per fraction ranging from 2 to 2.5 Gy were selected. Metastatic brain tumours were irradiated with stereotactic radiosurgery, with the prescription doses varying between 19 and 22 Gy in a single fraction.

All plans were created in the Tomotherapy treatment planning system v.5.1.1.6 (Accuray Inc., Sunnyvale, CA, USA) to be delivered by a Tomotherapy HD unit (Accuray Inc., Sunnyvale, CA, USA). A field width of 2.5 cm in dynamic jaw mode [16] was considered for head and neck and prostate cases and 1 cm for stereotactic brain plans. The initial modulation factor was set according to the planner's preferences and the adopted pitch values were based on published guidelines [4,5].

To evaluate plan deliverability, i.e., the agreement between planned and measured dose, pre-treatment QA verification results were retrospectively collected. All plans had been recalculated in the Tomotherapy phantom (Cheese phantom) and delivered with the couch out of the bore. Dosimetry Check software v.5.5 (LifeLine Software Inc., Austin, TX, USA) was used to reconstruct the measured dose distribution from the acquired sinogram [17,18]. 3D global gamma analysis was performed with 3% of maximum dose/3 mm distance-to-agreement criteria and 10% dose threshold (TH) for head and neck and prostate, and 3%/2 mm 10% TH for stereotactic brain plans. The passing rate acceptance limit was 95%. For the purpose of this work, more stringent criteria were also adopted, namely 3%/2 mm 10% TH for head and neck and prostate and 2%/2 mm 10% TH for stereotactic brain cases. For stereotactic brain plans, a Gafchromic EBT3 film (Ashland Inc., Covington, Kentucky, USA) was also used to assess the dose distribution in a coronal plane of the Cheese phantom. Films were scanned in a flatbed scanner Epson Expression 10000 XL (Seiko Epson Corporation, Japan) and a home-made software was utilized for film processing, applying triple-channel dosimetry [19]. Global gamma analysis was performed with a criterion of 3%/2 mm, in a dedicated Tomotherapy station. The passing rate acceptance limit was again 95%. Point dose measurements were also performed using an Exradin A1SL chamber (Standard Imaging, Middleton, WI, USA) placed in the same phantom at the centre of the emulated brain lesion. A difference between the planned and measured dose of ±3% was considered acceptable.

### 5.2.2 | Complexity metrics

For each plan, 14 complexity metrics, including some commonly evaluated parameters and novel metrics, have been computed from the planned sinogram using an in-house program developed in MATLAB R2017b (Mathworks, Natick, Massachusetts, USA). The planned sinogram is a two-dimensional matrix with information on the fraction of time that a given MLC leaf is open per projection, relative to its duration and it is saved in the DICOM RT plan.

The computed parameters included the actual modulation factor (MF). The MF is defined as the ratio between the maximum leaf open time and the mean of all non-zero leaf opening times. A

higher MF allows a larger range of beamlet intensities [6]. The mean LOT (in ms), the percentage of open leaves with an opening time below 100 ms (%LOT < 100 ms), 50 ms (%LOT < 50 ms), 30 ms (%LOT < 30 ms), and the percentage of leaves with an opening time close to the projection duration (%LOT > pT-20 ms) were also calculated to characterize the leaf open times distribution. As reported [2,5,7] the linear model assumed for the MLC leaf latency is violated for small leaf open times and LOTs approaching the projection duration, which may compromise plan deliverability. Therefore, plans with a lower mean LOT and a higher percentage of leaves with a short LOT and/or approaching the projection time were considered more complex.

The treatment time divided by the prescribed dose per fraction – TT/Gy – is here presented as a simple indicator of complexity. It depends somehow on the longitudinal extension of the target volume and on the pitch. Plans with a higher TT/Gy may be considered more complex.

To evaluate the modulation of the entire planned sinogram, i.e., the differences in leaf open times, three metrics are proposed. These are extensions to HT of indices previously defined for conventional IMRT techniques. The leaf open time variability (LOTV), adapted from [12], is calculated for each leaf that opens at least once during the treatment as:

$$LOTV_{leaf} = \frac{\sum_{k=1}^{N_{CP}-2} t_{max} - |fLOT(k, leaf) - fLOT(k+1, leaf)|}{(N_{CP} - 2) \times t_{max}} \quad (5.1)$$

where $t_{max}$ is the maximum LOT for that leaf across all control points (CP), $fLOT$ is the matrix corresponding to the planned sinogram and $N_{CP}$ the total number of control points that is equal to the number of projections +1. The plan LOTV corresponds to the average over all leaves. This index ranges from 0 to 1, being 1 when all leaves have the same opening time. The higher the LOTV, the lower the variations in leaf open times along the treatment, and therefore, the lower the plan modulation.

The Plan Time Sinogram Variation (PSTV), an adaptation of the plan intensity map modulation score proposed by Coselmon et al. [11], is computed for a given control point by summing the LOT differences in two directions:

$$\text{PSTV}_{CP} = \sum_{j=1}^{N_l - 1} |fLOT(CP, j) - fLOT(CP, j+1)| + |fLOT(CP, j) - fLOT(CP+1, j)| \quad (5.2)$$

where $N_l$ is the total number of MLC leaves (64). The PSTV for the plan is calculated as the mean of all $PSTV_{CP}$, as all the other indices defined hereafter. The larger the PSTV score the higher the MLC leaves open time variation, and therefore the plan modulation.

The Modulation Index (MI) [10] was here modified to quantify the leaf open time variations in the planned sinogram in the directions defined by the MLC leaves $(x)$, the projections/control points $(y)$ and the corresponding diagonals $(xy, yx)$. The number of LOT changes (Δt) between adjacent elements in the four directions that exceed a certain fraction $(f)$ of the standard deviation $(\sigma)$ of the entire planned sinogram was calculated $N(f; \Delta t > f\sigma)$ and hence the mean per projection:

$$Z_x(f) = \frac{1}{N_{CP} - 1} N_x(f; \Delta t_x > f\sigma) \quad (5.3.1)$$

$$Z_y(f) = \frac{1}{N_{CP} - 1} N_y\big(f; \Delta t_y > f\sigma\big) \quad (5.3.2)$$

$$Z_{xy}(f) = \frac{1}{N_{CP} - 1} N_{xy}\left(f; \Delta t_{xy} > f\sigma\right) \tag{5.3.3}$$

$$Z_{yx}(f) = \frac{1}{N_{CP} - 1} N_{yx}\left(f; \Delta t_{yx} > f\sigma\right) \tag{5.3.4}$$

where the total number of projections is equal to $N_{CP} - 1$ and $f = 0.01{:}\,0.01{:}\,2$ [10]. The total $Z(f)$ represents the spectrum of such changes in the entire sinogram and it is given by:

$$Z(f) = \frac{Z_x(f) + Z_y(f) + Z_{xy}(f) + Z_{yx}(f)}{4} \tag{5.4}$$

The modulation index corresponds to the area under the spectrum:

$$MI = \int_0^{2\sigma} Z(f)\, df \tag{5.5}$$

The larger the value of MI, the higher the plan modulation [10].

To assess the amount of tongue-and-groove effect in HT, two novel metrics are defined. The leaves with zero open neighbors score (L0NS) and the leaves with one open neighbor score (L1NS) are calculated for a given control point as:

$$L0NS_{CP}\,(\%) = \frac{Number\ of\ open\ leaves\ with\ zero\ open\ neighbors_{CP}}{Number\ of\ open\ leaves_{CP}} \times 100 \tag{5.6}$$

$$L1NS_{CP}\,(\%) = \frac{Number\ of\ open\ leaves\ with\ one\ open\ neighbor_{CP}}{Number\ of\ open\ leaves_{CP}} \times 100 \tag{5.7}$$

Due to the tongue-and-groove/penumbra blur effects, the primary fluence under a given MLC leaf varies according to the state of its neighbors. Such differences are taken into account in the TPS during the end-of-planning process, through a leaf-by-leaf basis correction [2,20]. The presented indices quantify the number of times that those corrections need to be applied, and eventually their accuracy.

The closed leaf score (CLS), adapted from [13], is computed per control point as the ratio of closed leaves to all MLC leaves (64):

$$CLS_{CP}\,(\%) = \frac{\sum_{j=1}^{N_l}(fLOT(CP, j) = 0)}{N_l} \times 100 \tag{5.8}$$

The CLS can vary between 0 and 100%, being 100% when all leaves are closed during the treatment. This index is partially related to the target volume. However, when the number of closed leaves per CP is high, it is assumed that a plan can be considered more complex due to the possible significant impact of mechanical errors and dose calculation uncertainties.

The percentage of closed leaves within the so-called treatment area, defined by the right most and left most open leaves in a given control point was also calculated ($CLS_{in}$). It gives an indication of the complexity of the irradiation pattern, due to the target volume irregularity and/or its proximity with critical structures. Thus, the higher the $CLS_{in}$ the greater the plan complexity.

### 5.2.3 | Statistical Analysis

Statistical analysis was performed in MATLAB. The mean and standard deviation of the 14 complexity metrics were calculated for the three groups of plans. As each plan is characterized by a vector of 14 features, as many as the considered complexity metrics, it is difficult to summarize, visualize and identify patterns in the data, namely differences between the groups. To reduce the dimensionality of the dataset and enable graphical representation of the distribution of plans, principal component analysis (PCA) was performed [21]. PCA is a multivariate technique widely used in dataset dimensionality reduction to increase interpretability while preserving most of the initial information [22-24]. For that, PCA finds a new set of uncorrelated variables (principal components, PCs) that result from linear combinations of the original ones and that successively maximize variance. The weight of each variable for every PC is known as its loading. The resulting number of PCs is equal to the number of original variables. To decide how many PCs to retain, a common approach consists in defining a cut-off (70%-90%) of the cumulative percentage of the total variance explained and consider the minimum number of PCs that exceed that cut-off. Another possibility consists in representing by the so-called scree plot, the variance associated with each PC vs PC number and base the decision on the analysis of the slope change between adjacent line segments [21].

In this study, all the 14 complexity metrics were considered for PCA analysis that corresponded to a data matrix X with 14 metrics (columns) for the 208 plans (rows). Those metrics had different units of measurement and numerical ranges, which affects the variance. To ensure that all variables would contribute equally to the analysis, data was standardized before performing PCA, such that all metrics had a mean of 0 and variance of 1 [25]. The PCA analysis output consisted of 14 principal components. To determine the number of PCs to keep for data representation, a cut-off of 70% of the total variance explained was adopted [21]. The scree plot for the HT complexity data is given in Supplementary Materials (Figure S5.1).

Still using PCA, after modifying some metrics such that all increased with increasing complexity, the methodology proposed by the authors in a previous work [25] was followed to compute a global plan complexity score (PCS). This score aims at characterizing and comparing the treatment plans through a single indicator and it is calculated as the weighted mean of the selected principal components:

$$Plan\ Complexity\ Score = \sum_{l=1}^{L} \frac{v_l}{v} \times PC_l \tag{5.9}$$

where $L$ is the minimum number of PCs corresponding to a cumulative percentage of the total variance explained higher than 70%, $v$ is the total variance explained by the retained PCs and $v_l$ the percentage of variance explained by $PC_l$.

The absolute value of the PCS may not be easy to interpret, as explained in Santos et al. [25]. Therefore, a normalized version of this score, $nPCS$, was calculated for a given plan $i$ within the set of plans as:

$$nPCS_i = \frac{PCS_i - min\ PCS}{max\ PCS - min\ PCS} \tag{5.10}$$

nPCS is 0 for the plan with the minimum PCS (min PCS) and 1 for the plan with the maximum PCS (max PCS). The higher the value of nPCS, the greater the plan complexity for the set of plans considered in the study.

The correlation between pre-treatment QA verification results, obtained for each group of plans, computed complexity metrics, and nPCS values was investigated using Spearman's rank correlation coefficients, for a significance level of 5%. Depending on the absolute value of $r_s$, the dependency was classified as: 0-0.19 "very weak", 0.20-0.39 "weak", 0.40-0.59 "moderate", 0.60-0.79 "strong" and 0.80-1 "very strong" [26].

## 5.3 | Results

### 5.3.1 | Treatment plans

Some of the TPS reported parameters for the considered groups of plans (head and neck, prostate and stereotactic brain) are summarized in Table 5.1. It can be seen that stereotactic brain plans have the longest gantry period and the highest number of gantry rotations, as well as the smallest pitch (0.100 for all plans), couch speed and couch travel. This is expected due to the high dose delivered in a single fraction (19-22 Gy) and the small target volume (5.9±5.1 cc, on average). Head and neck cases, on the other hand, present the highest pitch, fastest gantry period and couch speed. The couch travel gives an indication of the craneocaudal extension of the treatment region, being higher for the head and necks plans.

**Table 5.1 –** Summary of some conventional parameters (mean ± standard deviation) for the three groups of HT plans.

| | Head and Neck (n=105) | Prostate (n=51) | Stereotactic Brain (n=52) |
|---|---|---|---|
| Pitch | 0.408±0.032 | 0.348±0.042 | 0.100±0.000 |
| Gantry period (s) | 16.4±1.6 | 23.6±3.4 | 46.9±5.3 |
| Gantry rotations | 18.1±2.4 | 12.2±2.1 | 26.9±7.1 |
| Couch travel (cm) | 18.5±1.8 | 10.5±1.6 | 2.8±0.7 |
| Couch speed (cm/s) | 0.0629±0.0068 | 0.0376±0.0060 | 0.0023±0.0002 |

### 5.3.2 | Complexity metrics

Plans from various treatment sites were included in this study to appreciate the differences in terms of complexity between them based on the analysis of the planned sinogram. Figure 5.1 displays a representative example of a planned sinogram for each group. As Figure 5.1 illustrates, the head and neck plans typically have a more complex irradiation pattern. The prostate cases usually have a smaller number of projections whereas, the stereotactic brain plans are characterized by a low number of MLC open leaves per projection.

The visual differences in the planned sinogram complexity were quantified through the calculation of several metrics. Four pairs of those metrics are presented in Figure 5.2. Figure 5.2a) shows that the percentage of leaves with an open time close to the projection time (%LOT > pT-20 ms) tends to be higher for plans with a lower modulation factor (MF), especially when considering head and neck and prostate plans. Nevertheless, there are no obvious differences in the modulation factor between the groups. In Figure 5.2b) the percentage of leaves with an opening time below 100 ms (%LOT < 100 ms) is generally the highest in the head and neck plans and the lowest in stereotactic brain ones. No clear correlation was identified between the modulation factor within each group and the percentage of small LOT. Figure 5.2c) indicates that the modulation index allows a clear distinction between the groups of plans. LOTV did not change much among the groups meaning that the individual leaf open time variations along the treatment are generally quite smooth. The closed leaf score (CLS) was, on average, higher for the stereotactic brain plans as

could be anticipated due to the small volume of the metastatic lesions – Figure 5.2d). The percentage of leaves with one open neighbor evaluated by the L1NS is also the highest in this group, which is also somehow expected. As the number of open leaves per projection is reduced, the leaves in the extremities of the treatment area have more impact in its calculation.

A summary of the complexity metrics computed for the three groups of HT plans is presented in Table 5.2.

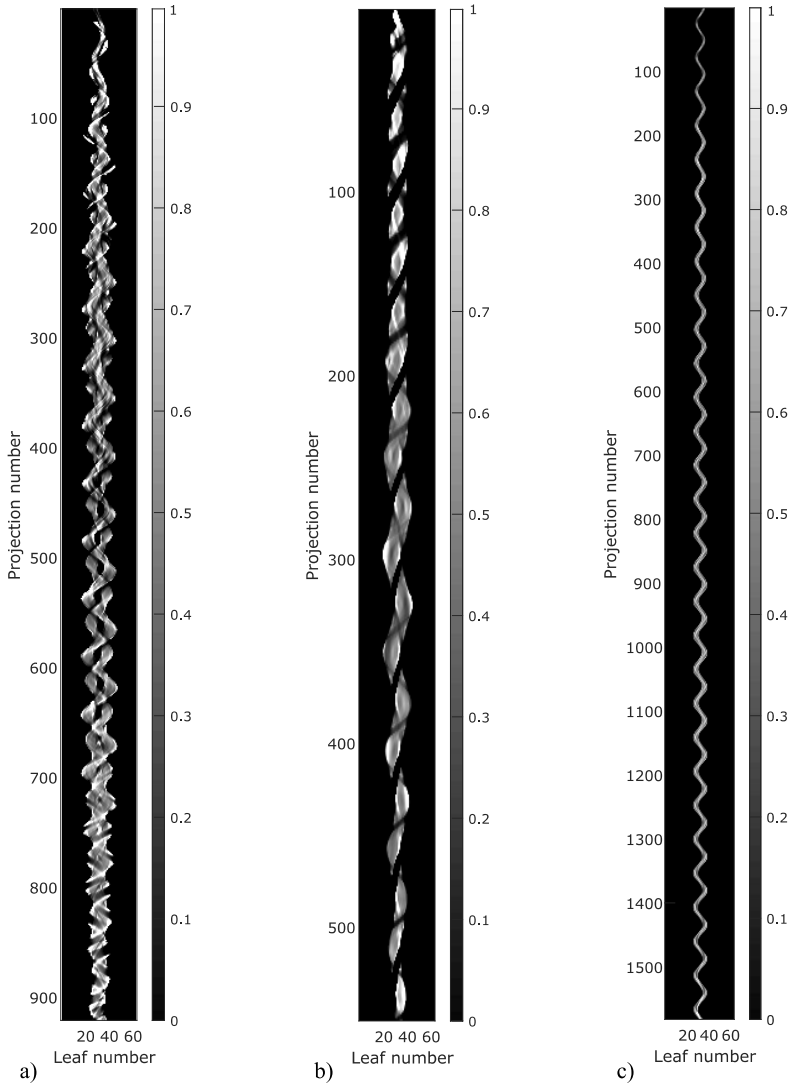


**Figure 5.1 –** Planned sinograms for a typical head and neck (a), prostate (b) and stereotactic brain (c) case. Each image element (row, column) represents the fraction of time (0-1) that a given MLC leaf is open relative to the projection duration.

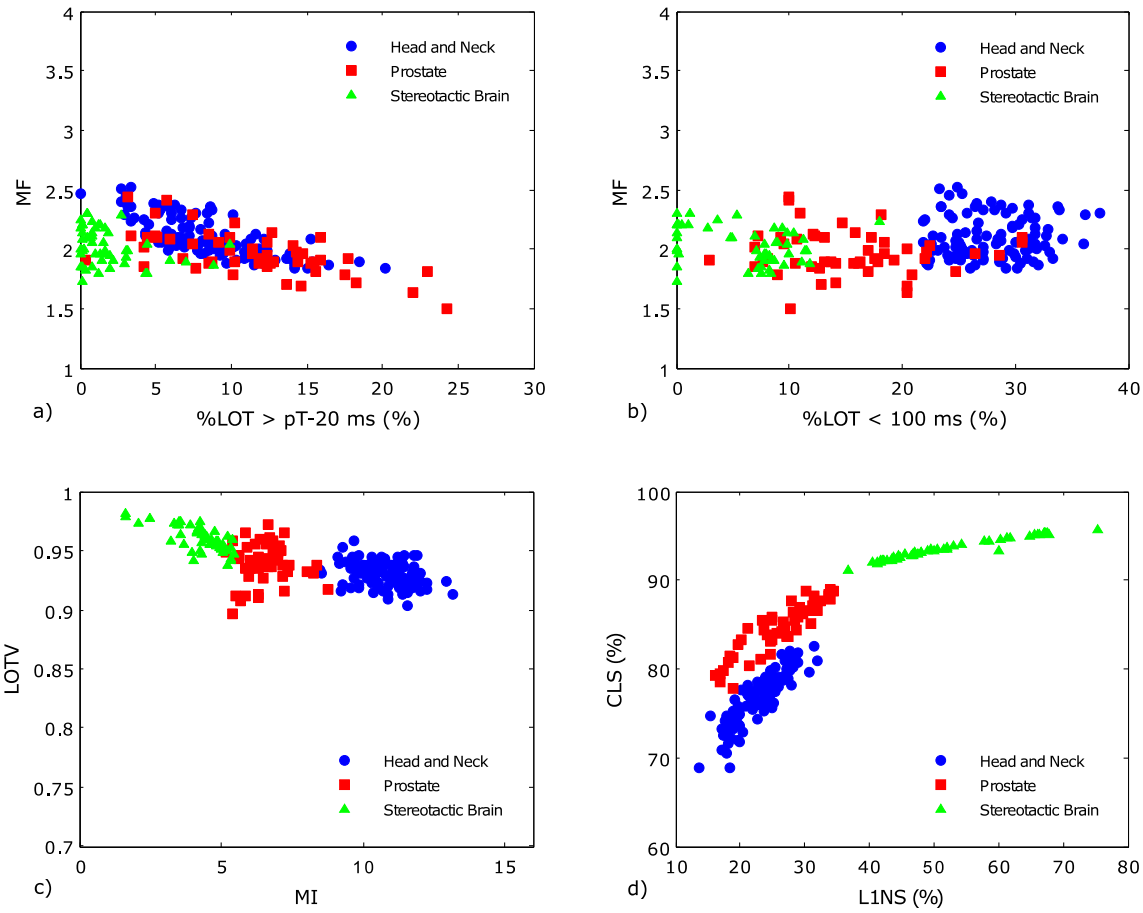


**Figure 5.2 –** Representation of four pairs of complexity metrics for the three treatment sites.

**Table 5.2 –** Average of the complexity metrics (mean ± standard deviation) for the three groups of HT plans.

| | Head and Neck (n=105) | Prostate (n=51) | Stereotactic Brain (n=52) |
|---|---|---|---|
| Modulation Factor (MF) | 2.096±0.173 | 1.966±0.184 | 2.027±0.151 |
| TT/Gy (s/Gy) | 141.6±20.6 | 133.2±28.5 | 60.1±12.9 |
| Mean LOT (ms) | 153.2±9.4 | 234.7±26.7 | 452.7±35.1 |
| %LOT < 100 ms (%) | 27.9±3.6 | 15.5±6.3 | 5.9±4.5 |
| %LOT < 50 ms (%) | 9.7±1.8 | 5.8±2.7 | 2.9±2.4 |
| %LOT < 30 ms (%) | 3.7±0.8 | 2.3±1.0 | 1.3±1.1 |
| %LOT > pT-20 ms (%) | 8.7±3.7 | 11.2±5.2 | 1.8±2.2 |
| LOTV | 0.931±0.010 | 0.939±0.016 | 0.960±0.010 |
| PSTV | 5.4±0.7 | 3.4±0.7 | 1.8±0.5 |
| Modulation Index (MI) | 10.7±0.9 | 6.6±0.8 | 4.3±0.9 |
| CLS (%) | 76.7±3.0 | 84.6±2.9 | 93.4±1.2 |
| $CLS_{in}$ (%) | 10.3±4.5 | 5.4±4.0 | 0.2±1.5 |
| L0NS (%) | 0.9±0.5 | 1.4±1.2 | 0.0±0.1 |
| L1NS (%) | 22.8±3.8 | 26.2±5.2 | 51.7 ±9.7 |

From the computed metrics, PCA was performed to reduce the dataset dimensionality. In PCA, the first two principal components together explained 76.5% of the total variance, 65.2% and 11.3%, respectively, which is above the predefined cut-off (70%). Therefore, the resulting two-dimensional representation of the data can be considered a good approximation of the original scatter plot in 14 dimensions.

Figure 5.3 presents the PC2 vs PC1 for the HT data, where each point corresponds to a plan (primary axis). The metrics are also displayed, as vectors, where the x-component represents the weight of that variable for the PC1 and y-component the weight for the PC2 (secondary axis). This kind of representation provides information on the correlation between the 14 complexity metrics for the entire set of plans. If two variables have a large positive association, the corresponding vectors are close together; if it is weak, the angle between them is around 90° and when negative and strong the angle is close to 180°. For instance, metrics that assess the percentage of leaves with small leaf open times (%LOT < 100 ms, %LOT < 50 ms and %LOT < 30 ms) present a very strong dependency, $r_s$>0.8. These indices are in turn, negatively correlated with the mean LOT, which indicates, as expected, that plans with a smaller mean LOT have a higher percentage of short leaf open times. None of these parameters is associated with the modulation factor (MF). The %LOT > pT-20 ms, on the contrary, confirmed to be negatively correlated with the modulation factor. To further illustrate this inverse relationship, the TPS LOT histograms of three prostate cases are provided in the Supplementary Materials – Figure S5.2. Regarding the indices that evaluate the variations of leaf open times, the relationship between the modulation index (MI) and the LOTV is negative, $r_s$=-0.691, whereas it is positive with the PSTV. These indices (MI, LOTV and PSTV) are in turn, strongly associated with the $CLS_{in}$ that evaluates the percentage of closed leaves within the treatment area. Finally, the L1NS presents a very strong relationship with the CLS, $r_s$=0.891, as referred above.

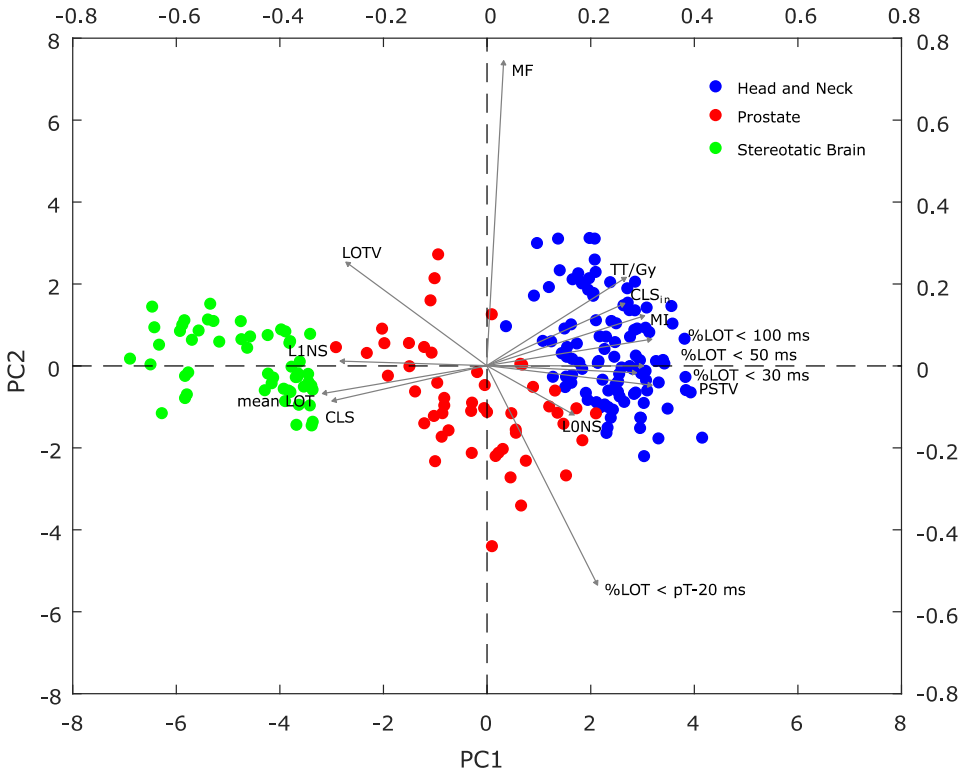


**Figure 5.3 –** Biplot for the HT treatment plans data. Each data point corresponds to a plan and the weights of each variable are represented by a vector.

From the distribution of the data points, three clusters can be fairly identified, corresponding to the considered groups of plans, indicating that they have generally different levels of complexity. Further analysis of the relative position of the data points and the variables can explain groups' separation. PC1, represented by the horizontal axis, accounts for 65.2% of the total variation of the dataset and it is indeed the PC that more clearly separates the groups of plans. Plans on the left side of the graph, having large negative values for PC1, tend to have higher values for the variables with larger negative loadings for that PC and lower values for the variables with large positive loadings. Therefore, stereotactic brain plans, on the left side of the figure, must have the highest mean LOT, percentage of closed leaves per projection (CLS), L1NS and LOTV. Head and neck plans, on the right side, tend to have the highest percentage of small leaf open times (%LOT < 100 ms, %LOT < 50 ms and %LOT < 30 ms), MI, PSTV, and percentage of closed leaves within the treatment area ($CLS_{in}$). These findings are corroborated by the information in Table 5.2. Prostate plans are mostly in between the other two groups, except some cases that are overlapped with the head and neck one. Looking now at the vertical axis, PC2, the modulation factor (MF), and the %LOT > pT-20 ms are the variables that mostly contribute to this PC. But while plans on the top of the figure tend to have a higher modulation factor, those towards the bottom have a higher percentage of leaves with an open time close to the projection duration.

To summarize the information provided in the biplot and compare the global complexity inter- and intra- group of plans, the normalized plan complexity score (nPCS) was calculated – Figure 5.4. As the median line of each boxplot in Figure 5.4 lies outside the other boxes, the groups of plans can be considered different in terms of global complexity. The head and neck plans are generally the most complex, followed by the prostate and the stereotactic brain ones. Stereotactic brain plans have the highest interquartile range whereas the prostate group presented the largest spread of nPCS, as indicated by the extreme values of the corresponding boxplot.

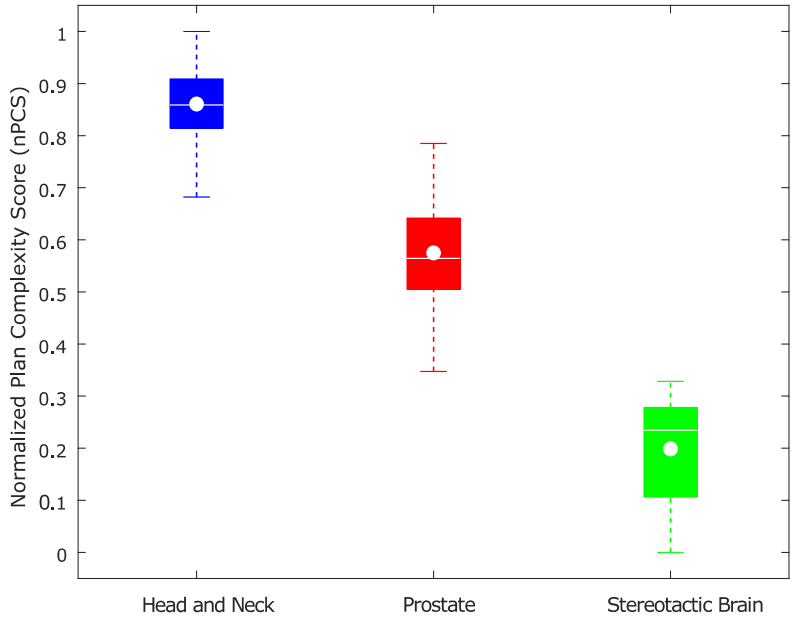


**Figure 5.4 –** Boxplots summarizing the normalized plan complexity score (nPCS) obtained for the three groups of plans. On each box, the bottom and top edges indicate the 25th and 75th percentiles, respectively – interquartile range. The median value is represented by a white horizontal line and the mean value by a white 'o'. The whiskers go down to the minimum and up to the maximum data values within 1.5 times the interquartile range.

### 5.3.3 | Correlation between the complexity metrics and pre-treatment QA results

Pre-treatment verification results were acceptable for all plans, with no significant differences between the groups. The 3D global gamma passing rate for the head and neck cases was on average 98.6±1.0% and 97.6±1.9% for the prostate plans, considering a 3%/3 mm, 10% TH criteria. The average results for the stereotactic brain group were: 98.7±2.5% – 3D global gamma passing rate (3%/2 mm, 10% TH), 0.1±1.1% – IC % difference, and 98.5±1.5% – film gamma passing rate (3%/2 mm).

The correlation between each complexity metric/global complexity score and the results of pre-treatment QA verification obtained for each group of plans was investigated through the calculation of the Spearman's rank correlation coefficients. For the head and neck and prostate cases, no significant correlations have been identified, neither when considering each metric individually nor the nPCS (Table S5.1 in Supplementary Materials). For the stereotactic brain group, the obtained Spearman's rank correlation coefficients and corresponding p-values are presented in Table 5.3. Some moderate and strong dependencies have been obtained for this group of plans. The correlations tended to be stronger when more stringent analysis criteria were adopted for 3D global gamma analysis. Nevertheless, ionization chamber results were not related with any of the computed metrics.

**Table 5.3 –** Spearman's correlation coefficients, $r_s$, and corresponding p-values (within brackets) between 3D gamma passing rates with various criteria, ionization chamber percent difference (IC %diff), film results and the complexity metrics/nPCS for the stereotactic brain plans. Correlations were considered statistically significant for a p-value < 0.05. Values in bold correspond to significant moderate or strong correlations.

| | 3D global gamma analysis | | IC %diff | Film |
|---|---|---|---|---|
| | 3%/2 mm 10% TH | 2%/2 mm 10% TH | | 3%/2 mm |
| Modulation Factor | 0.148 (0.294) | 0.320 (0.021) | -0.076 (0.594) | 0.385 (0.008) |
| TT/Gy | -0.346 (0.012) | **-0.659 (0.000)** | -0.185 (0.189) | **-0.463 (0.001)** |
| Mean LOT | -0.035 (0.804) | 0.327 (0.018) | 0.138 (0.330) | 0.203 (0.176) |
| %LOT < 100 ms | -0.144 (0.307) | -0.373 (0.006) | 0.056 (0.692) | -0.329 (0.025) |
| %LOT < 50 ms | -0.168 (0.233) | -0.353 (0.010) | 0.080 (0.575) | -0.363 (0.013) |
| %LOT < 30 ms | -0.215 (0.127) | -0.387 (0.005) | 0.119 (0.399) | -0.393 (0.007) |
| %LOT > pT-20 ms | -0.303 (0.029) | **-0.463 (0.001)** | -0.082 (0.565) | **-0.440 (0.002)** |
| LOTV | 0.278 (0.046) | 0.349 (0.012) | -0.041 (0.772) | 0.303 (0.040) |
| PSTV | -0.291 (0.036) | **-0.436 (0.001)** | 0.111 (0.432) | -0.337 (0.022) |
| Modulation Index (MI) | -0.231 (0.100) | -0.329 (0.018) | 0.091 (0.522) | -0.023 (0.878) |
| CLS | 0.378 (0.006) | **0.655 (0.000)** | 0.103 (0.466) | 0.353 (0.016) |
| $CLS_{in}$ | -0.011 (0.937) | 0.030 (0.832) | -0.039 (0.786) | 0.018 (0.908) |
| L0NS | 0.029 (0.839) | 0.041 (0.775) | -0.009 (0.952) | -0.064 (0.671) |
| L1NS | 0.395 (0.004) | **0.678 (0.000)** | 0.089 (0.532 | 0.329 (0.025) |
| nPCS | -0.253 (0.070) | **-0.434 (0.001)** | 0.024 (0.867) | **-0.442 (0.002)** |

Plans with a higher total treatment time per Gy (TT/Gy) were significantly associated with poorer verification results, such that $r_s$=-0.463, p-val=0.001 for film. The negative relationships observed for %LOT > pT-20 ms also suggest that a high percentage of leaves with an opening time close to the projection duration may compromise plans deliverability. As for the metrics proposed to evaluate the variation in leaf open times, a significant dependency was observed between the PSTV and the 3D global gamma passing rates, $r_s$=-0.436, p-val=0.001. The positive and strong

dependency for the L1NS index and the more stringent 3D gamma analysis may indicate that the amount of tongue-and-groove does not have any influence on treatment deliverability. The same rationale applies to the CLS index that evaluates the percentage of leaves closed per control point. Lastly, the normalized plan complexity score showed a moderate association with the verification results ($r_s$>0.4, p-val=0.000).

The correlation between some of the TPS reported plan parameters, namely, the pitch, the gantry period, the number of gantry rotations, the couch travel and the couch speed and the pre-treatment QA results was also investigated (Table S5.2 in Supplementary Materials). But once again, no significant dependencies have been identified for the head and neck and the prostate groups. Nevertheless for the stereotactic brain plans, all parameters, except pitch (0.100 for all plans) were significantly associated with the film gamma passing rates ($r_s$>0.4, p-val<0.05). The correlation was negative with the number of gantry rotations, couch speed, and couch travel and positive with the gantry period.

## 5.4 | Discussion

In this study, the complexity of a set of helical tomotherapy plans from head and neck, prostate and brain treatment sites was characterized and compared. To our knowledge, this is the first work that attempts to comprehensively extend the current discussion on plan complexity to HT, through both the definition of individual metrics and the calculation of a global complexity score.

In total, 14 complexity metrics were computed, including some commonly evaluated parameters and novel indices with the aim of assessing different features of the treatment plans. Due to the high number of metrics, an approach based on principal component analysis was followed to explore the correlation between them to find a subset of the most representative components. The visual representation of the data in two-dimensions allowed the identification of similarities and differences between the considered treatment sites.

The head and neck plans were found to be the most complex for almost all the computed complexity indicators. Plans from this group had, for instance, the highest percentage of leaves with short opening times and the lowest mean LOT. Among the considered groups, the head and neck plans had the fastest gantry period which confirms the known influence of this parameter in the amount of leaves with small LOTs [2,27]. The score that evaluates the percentage of closed leaves within the treatment area ($CLS_{in}$) indicated that the head and neck plans had also the most complex irradiation patterns which was confirmed by the calculated modulation indices, LOTV, PSTV and MI. The prostate plans had the highest percentage of leaves with an opening time approaching the projection duration (%LOT > pT-20 ms). The %LOT > pT-20 ms has shown to be inversely correlated with the modulation factor, which is in accordance with the findings by Binny et al. [27] and Sevillano et al. [7]. The stereotactic brain plans had the highest percentage of closed leaves per projection, with just about five open leaves, on average, as well as the largest L1NS, which quantifies the amount of tongue-and-groove effect.

To summarize the information given by all complexity metrics, a global plan complexity score (nPCS) was calculated, following the methodology proposed by Santos et al. [25] in the context of a national IMRT audit [28]. The nPCS combines the multiple metrics into a single numerical score, allowing for the comparison of the relative complexity of the entire set of plans. Based on the nPCS values, the head and neck plans confirmed to be the most complex, followed by the prostate and the stereotactic brain ones. A higher complexity variability was observed for the prostate cases, presenting the larger range of nPCS values. This can be explained, in part, by the inclusion of patients with femoral prosthesis, demanding an adaptation of the typical planning strategy to reduce

dose calculation uncertainties. The number of planners was also higher for the prostate group (6) than for the head and neck (3) and stereotactic brain (2) groups, being the adopted planning strategies largely dependent on the planners' skills and approaches [29]. Yet, the reported differences in plan complexity intra- and inter-group may also be due to factors such as variations in patient anatomy, PTV shape and/or volume, and in the dose constraints that may differ from one clinician to another. Plans complexity is also partially determined by the optimization algorithm that works like a black box. In HT, only the initial modulation factor, pitch and field width are set, and the actual modulation factor, number of projections and leaf open times result from the optimization process [9].

The plans deliverability was weakly correlated with both the computed complexity metrics and the usual plan parameters for head and neck and prostate groups. Regarding the stereotactic brain plans, some moderate dependencies have been identified. Stereotactic brain plans with a higher total treatment time per Gy (TT/Gy) and/or with a higher percentage of leaves with opening times approaching the projection duration (%LOT > pT-20 ms) were significantly associated with a poorer agreement (although within tolerance) between planned and measured dose. Plan parameters like the couch speed and the number of gantry rotations were also inversely correlated with the pre-treatment QA results. These findings may be useful to establish clinical guidelines for planning of stereotactic brain cases at our institution. Accordingly, planners should carefully evaluate the TPS LOT distribution during the planning phase and aim to achieve a %LOT > pT-20 ms less than the reported mean. The TT/Gy can also be easily assessed and it should be compared with the values obtained for similar clinical cases belonging to this group. The same rationale applies to the planning parameters.

All the 208 plans considered in this study had pre-treatment QA results within the established tolerances, which is perhaps the main reason for the reported lack of significant correlations, mainly for the head and neck and prostate groups. A pool of plans with poorer verification results would perhaps allow for a clearer insight into the relationship between complexity and deliverability. Still, there are some factors that can mask the effect of complexity on delivery verification results. The first one is related to the pre-treatment QA methods used. For the head and neck and prostate cases, the verification was based on 3D dose reconstruction using the signal recorded by the exit detector during delivery with the couch out of the bore. This methodology even having been deeply tested against film and ionization chamber measurements [18], may raise some questions related with the detector resolution, the dose calculation algorithm used (pencil beam) and even the sensitivity of the global gamma analysis [30,31]. TPS dose calculation (beam modeling, MLC modeling, commissioning inaccuracies) and MLC performance are also some potential sources of errors.

The impact of HT plan parameters on the pre-treatment QA results has already been investigated by other authors [2,8,9]. Westerly et al. [2], for instance, found that plans with a mean LOT below 100 ms, i.e. using predominantly small leaf open times, were more likely to present significant deviations (>3%) between the calculated and the measured doses, in an analysis of six plans treated in a Tomotherapy Hi-Art II system. Bresciani et al. [8] have taken a pool of 384 HT plans to investigate if there were any treatment sites and/or plan parameters more likely to lead to poorer verification results. A wide variety of treatment sites were included, such as gynecological, head and neck, breast, lung, rectum, but no significant differences were noticed. As for the plan parameters, the correlations were rather weak. Binny et al. [9] in a retrospective study with 28 head and neck, 19 pelvic and 23 brain HT plans, found no correlations between the TPS reported parameters, such as the gantry period and the percentage of leaf open times less than 100 ms, and the pre-treatment QA results.

The characterization of HT plans complexity is particularly challenging. Due to the system particularities, defining what is more or less complex and which features to assess is not trivial. As

most parameters reported by the TPS are intimately linked with each other and depend on the target volume and dose per fraction, it was decided not to include them in this study as complexity metrics. Also, new indices were defined that evaluate different aspects of the plans that may directly or indirectly contribute to increased uncertainties in dose calculation and delivery. These indicators demonstrated to be effective in quantifying the complexity of the plans. The reported differences inter- and intra- group suggest that it may be appropriate to define site-specific recommendations to guide the planning and the QA processes. The values of both planning parameters and complexity metrics may be generally adopted as reference levels at our institution, as the pre-treatment QA results of the plans included in this study were all clinically acceptable.

Some of the proposed complexity metrics, namely the modulation factor, TT/Gy, mean LOT, %LOT < 100 ms, %LOT < 50 ms, %LOT < 30 ms and the %LOT > pT-20 ms can be directly assessed during the planning phase. Their comparison against local reference values may be used as a guide for planning and eventually contribute to harmonize local planning strategies. Planners should, along this line, carefully evaluate the percentage of leaves with a LOT approaching the projection duration and keep it as close as possible to the reference value for each site. Namely, in prostate cases where some atypically high values (>20%) have been identified and in stereotactic brain plans where a significantly correlation with the pre-treatment QA results has been noticed. The indices that evaluate the plan sinogram modulation and the amount of tongue-and-groove effect, on the contrary, can only be calculated after planning. Their calculation would not necessarily increase the time allocated to the QA process. Again, the comparison of the obtained values for a new plan with the reference ones may be useful to flag plans with a complexity higher than usual, which would be subject to a more rigorous QA.

## 5.5 | Conclusion

In this study, the complexity of HT plans from different treatment sites was characterized and compared through the calculation of a set of metrics that evaluate multiple features of the planned sinograms. A statistical approach based on principal component analysis was followed to simplify data interpretation, allowing to explore the correlations among the proposed indices and quantify the differences in complexity between the studied groups of plans. Generally, head and neck plans were found to be the most complex for almost all metrics, which was confirmed by the computed global plan complexity score. The prostate plans had the highest complexity variability, which can be a result of a wider range of planning approaches.

The presented characterization of the differences inter- and intra- group of treatment sites may be useful to guide the treatment planning and the QA processes eventually reducing uncertainties and harmonizing local planning strategies.

## 5.6 | Acknowledgments

This work was supported by FCT, the Portuguese Foundation for Science and Technology, through a PhD scholarship, reference SFRH/BD/118929/2016 and the project POCI-01-0145-FEDER-028030.

## 5.7 | Conflict of interest

None.

## 5.8 | References


1. Jeraj R, Mackie TR, Balog J, et al. Radiation characteristics of helical tomotherapy. Med Phys. 2004;31(2):396-404. doi:10.1118/1.1639148
2. Westerly DC, Soisson E, Chen Q, et al. Treatment planning to improve delivery accuracy and patient throughput in helical tomotherapy. Int J Radiat Oncol Biol Phys. 2009;74(4):1290-1297. doi:10.1016/j.ijrobp.2009.02.004
3. Boyd R, Jeong K, Tomé WA. Determining efficient helical IMRT modulation factor from the MLC leaf-open time distribution on precision treatment planning system. J Appl Clin Med Phys. 2019;20(5):64-74. doi:10.1002/acm2.12581
4. Kissick MW, Fenwick J, James JA, et al. The helical tomotherapy thread effect. Med Phys. 2005;32(5):1414-1423. doi:10.1118/1.1896453
5. Chen M, Chen Y, Chen Q, Lu W. Theoretical analysis of the thread effect in helical TomoTherapy. Med Phys. 2011;38(11):5945-5960. doi:10.1118/1.3644842
6. Shimizu H, Sasaki K, Tachibana H, et al. Analysis of modulation factor to shorten the delivery time in helical tomotherapy. J Appl Clin Med Phys. 2017;18(3):83-87. doi:10.1002/acm2.12075
7. Sevillano D, Minguez C, Sanchez A, Sanchez-Reyes A. Measurement and correction of leaf open times in helical tomotherapy. Med Phys. 2012;39(11):6972-6980. doi:10.1118/1.4762565
8. Bresciani S, Miranti A, Di Dia A, et al. A pre-treatment quality assurance survey on 384 patients treated with helical intensity-modulated radiotherapy. Radiother Oncol. 2016;118(3):574-576. doi:10.1016/j.radonc.2015.12.021
9. Binny D, Lancaster CM, Byrne M, Kairn T, Trapp J V, Crowe SB. Tomotherapy treatment site specific planning using statistical process control. Phys Medica. 2018;53:32-39. doi:10.1016/j.ejmp.2018.08.003
10. Webb S. Use of a quantitative index of beam modulation to characterize dose conformality: illustration by a comparison of full beamlet IMRT, few-segment IMRT (fsIMRT) and conformal unmodulated radiotherapy. Phys Med Biol. 2003;48:2051-2062. doi:10.1088/0031-9155/48/14/301
11. Coselmon MM, Moran JM, Radawski JD, Fraass BA. Improving IMRT delivery efficiency using intensity limits during inverse planning. Med Phys. 2005;32(5):1234-1245. doi:10.1118/1.1895545
12. McNiven AL, Sharpe MB, Purdie TG. A new metric for assessing IMRT modulation complexity and plan deliverability. Med Phys. 2010;37(2):505-515. doi:10.1118/1.3276775
13. Crowe SB, Kairn T, Kenny J, et al. Treatment plan complexity metrics for predicting IMRT pre-treatment quality assurance results. Australas Phys Eng Sci Med. 2014;37(3):475-482. doi:10.1007/s13246-014-0274-9
14. Du W, Cho SH, Zhang X, Hoffman KE, Kudchadker RJ. Quantification of beam complexity in intensity-modulated radiation therapy treatment plans. Med Phys. 2014;41(2):21716. doi:10.1118/1.4861821
15. Park JM, Park S-Y, Kim H, Kim JH, Carlson J, Ye S-J. Modulation indices for volumetric modulated arc therapy. Phys Med Biol. 2014;59(23):7315-7340. doi:10.1088/0031-9155/59/23/7315
16. Sterzing F, Uhl M, Hauswald H, et al. Dynamic Jaws and Dynamic Couch in Helical Tomotherapy. Int J Radiat Oncol Biol Phys. 2010;76(4):1266-1273. doi:10.1016/j.ijrobp.2009.07.1686
17. Narayanasamy G, Zalman T, Ha CS, Papanikolaou N, Stathakis S. Evaluation of Dosimetry Check software for IMRT patient-specific quality assurance. J Appl Clin Med Phys. 2015;16(3):5427. doi:10.1120/jacmp.v16i3.5427
18. Santos T, Ventura T, Mateus J, Capela M, Lopes MC. EP-1751: A comparison of tools for Delivery Quality Assurance in TomoTherapy. Radiother Oncol. 2017;123:S964. doi:10.1016/S0167-8140(17)32114-X
19. Ventura T, Lopes MC, Ferreira BC. EP-1517: Multichannel absolute film dosimetry for EBT3 GafChromic films. Radiother Oncol. 2014;111:S169. doi:10.1016/S0167-8140(15)31635-2
20. Yuan J, Rong Y, Chen Q. A virtual source model for Monte Carlo simulation of helical tomotherapy. J Appl Clin Med Phys. 2015;16(1):4992. doi:10.1120/jacmp.v16i1.4992

21. Jolliffe IT. “Principal component analysis” 2nd ed. New York, NY: Springer; 2002
22. Gurney-Champion OJ, Collins DJ, Wetscherek A, et al. Principal component analysis fosr fast and model-free denoising of multi b-value diffusion-weighted MR images. Phys Med Biol. 2019;64(10):105015. doi:10.1088/1361-6560/ab1786
23. Tsiamas P, Bagher-Ebadian H, Siddiqui F, et al. Principal component analysis modeling of Head-and-Neck anatomy using daily Cone Beam-CT images. Med Phys. 2018;45(12):5366-5375. doi:10.1002/mp.13233
24. Benestad R, Parding K, Dobler A, Mezghani A. A strategy to effectively make use of large volumes of climate data for climate change adaptation. Clim Serv. 2017;6:48-54. doi:10.1016/j.cliser.2017.06.013
25. Santos T, Ventura T, Lopes MC. Evaluation of the complexity of treatment plans from a national IMRT/VMAT audit – Towards a plan complexity score. Phys Medica. 2020;70. doi: 10.1016/j.ejmp.2020.01.015
26. Weir J. Spearman’s Rank Correlation – Introduction. http://www.statstutor.ac.uk/types/teach-yourself/spearmans-correlation-coefficient/. Acessed April 3, 2020
27. Binny D, Lancaster CM, Harris S, Sylvander SR. Effects of changing modulation and pitch parameters on tomotherapy delivery quality assurance plans. J Appl Clin Med Phys. 2015;16(5):87-105. doi:10.1120/jacmp.v16i5.5282
28. Santos T, Lopes MC, Gershkevitsh E, et al. IMRT national audit in Portugal. Phys Medica. 2019;65. doi:10.1016/j.ejmp.2019.08.013
29. Nelms BE, Robinson G, Markham J, et al. Variation in external beam treatment plan quality: An inter-institutional study of planners and planning systems. Pract Radiat Oncol. 2012;2(4):296-305. doi:10.1016/j.prro.2011.11.012
30. Gordon J, Siebers J. Addressing a gap in current IMRT quality assurance. Int J Radiat Oncol Biol Phys. 2013;87(1):20-21. doi:10.1016/j.ijrobp.2013.03.030
31. Nelms BE, Zhen H, Tome WA. Per-beam, planar IMRT QA passing rates do not predict clinically relevant patient dose errors. Med Phys. 2011;38(2):1037-1044. doi:10.1118/1.3544657

## 5.9 | Supplementary materials

**Table S5.1 –** Spearman's correlation coefficients, $r_s$, and corresponding p-values (within brackets) between 3D gamma passing rates with various criteria and the complexity metrics/nPCS for the head and neck and prostate plans.

| | **Head and Neck** | | **Prostate** | |
|---|---|---|---|---|
| | 3%/3 mm 10% TH | 3%/2 mm 10% TH | 3%/3 mm 10% TH | 3%/2 mm 10% TH |
| Modulation Factor (MF) | 0.110 (0.263) | 0.070 (0.481) | -0.159 (0.265) | -0.189 (0.184) |
| TT/Gy | 0.221 (0.023) | 0.249 (0.010) | -0.367 (0.008) | -0.332 (0.017) |
| Mean LOT | -0.263 (0.007) | -0.264 (0.006) | 0.119 (0.406) | 0.100 (0.486) |
| %LOT < 100 ms | -0.202 (0.039) | -0.201 (0.040) | -0.332 (0.017) | -0.260 (0.066) |
| %LOT < 50 ms | -0.292 (0.003) | -0.287 (0.003) | -0.222 (0.117) | -0.162 (0.255) |
| %LOT < 30 ms | -0.291 (0.003) | -0.300 (0.002) | -0.127 (0.373) | -0.068 (0.636) |
| %LOT > pT-20 ms | -0.285 (0.003) | -0.256 (0.008) | -0.194 (0.173) | -0.132 (0.356) |
| LOTV | 0.351 (0.000) | 0.339 (0.000) | 0.315 (0.000) | 0.357 (0.001) |
| PSTV | -0.349 (0.000) | -0.338 (0.000) | -0.326 (0.020) | -0.248 (0.080) |
| Modulation Index (MI) | -0.141 (0.150) | -0.148 (0.132) | 0.029 (0.842) | 0.047 (0.746) |
| CLS | 0.144 (0.143) | 0.160 (0.103) | -0.291 (0.038) | -0.266 (0.059) |
| $CLS_{in}$ | -0.109 (0.270) | -0.140 (0.156) | -0.399 (0.004) | -0.325 (0.020) |
| L0NS | -0.048 (0.624) | -0.025 (0.799) | -0.357 (0.001) | -0.383 (0.006) |
| L1NS | -0.038 (0.703) | -0.039 (0.694) | -0.367 (0.008) | -0.318 (0.023) |
| **nPCS** | -0.325 (0.001) | -0.328 (0.001) | -0.369 (0.001) | -0.385 (0.005) |

**Table S5.2 –** Spearman's correlation coefficients, $r_s$, and corresponding p-values (within brackets) TPS reported parameters and the pre-treatment QA results for the head and neck, prostate and SRS plans. Values in bold correspond to significant moderate or strong correlations.

| | pitch | Gantry period | Gantry rotations | Couch travel | Couch speed |
|---|---|---|---|---|---|
| Head and Neck | | | | | |
| 3D global gamma analysis | | | | | |
| 3%/3 mm 10% TH | -0.136 (0.165) | -0.101 (0.303) | 0.322 (0.001) | -0.294 (0.002) | -0.077 (0.433) |
| 3%/2 mm 10% TH | -0.126 (0.200) | -0.137 (0.165) | 0.377 (0.000) | 0.373 (0.000) | -0.035 (0.725) |
| Prostate | | | | | |
| 3D global gamma analysis | | | | | |
| 3%/3 mm 10% TH | 0.199 (0.160) | -0.057 (0.690) | -0.106 (0.458) | 0.178 (0.210) | 0.312 (0.026) |
| 3%/2 mm 10% TH | 0.175 (0.220) | -0.100 (0.483) | -0.057 (0.692) | 0.209 (0.141) | 0.343 (0.014) |
| Stereotactic brain | | | | | |
| 3D global gamma analysis | | | | | |
| 3%/2 mm 10% TH | - | 0.079 (0.580) | -0.342 (0.013) | -0.347 (0.012) | -0.079 (0.580) |
| 2%/2 mm 10% TH | - | **0.410 (0.003)** | **-0.657 (0.000)** | **-0.660 (0.000)** | **-0.410 (0.003)** |
| Film 3%/2 mm | - | 0.068 (0.632) | -0.182 (0.196) | -0.186 (0.187) | -0.068 (0.632) |
| IC %diff | - | 0.397 (0.006) | **-0.489 (0.001)** | **-0.486 (0.001)** | -0.397 (0.006) |

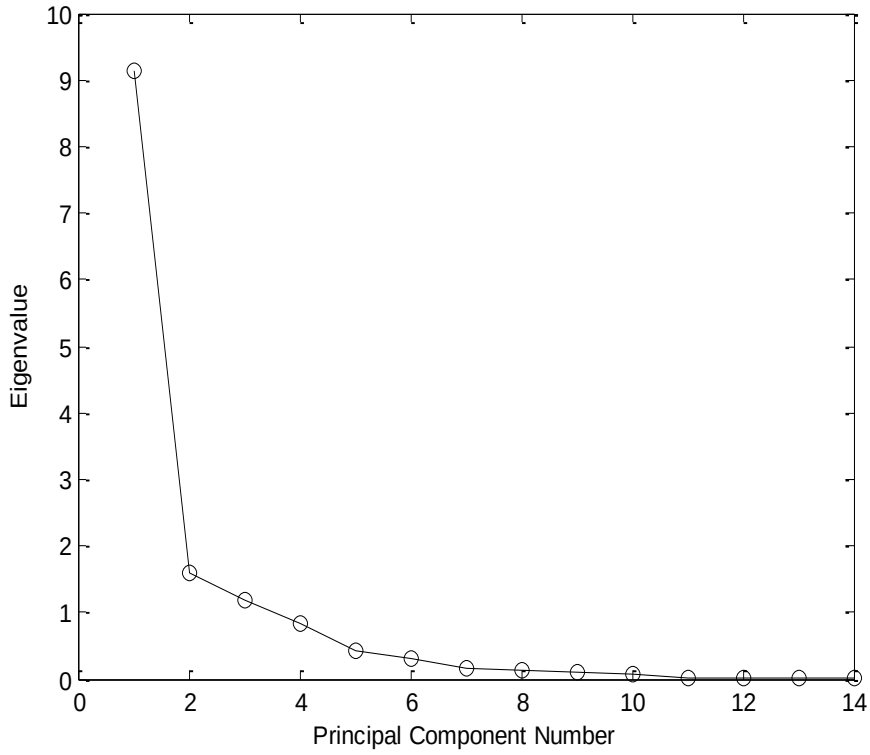


**Figure S5.1 –** Scree plot of the principal components for the HT treatment plans data. The eigenvalues give the variance explained by each PC.

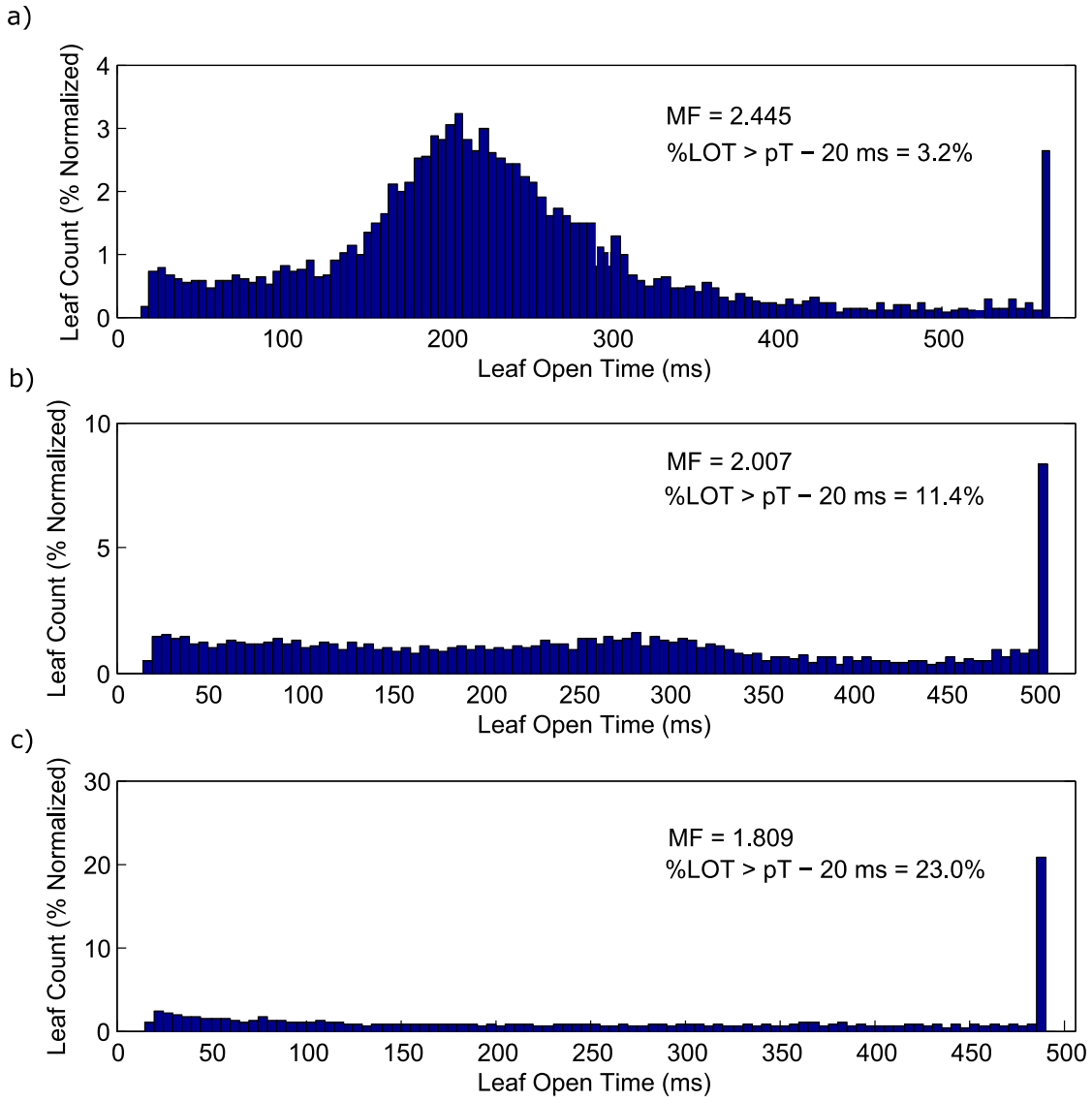


**Figure S5.2 –** TPS leaf open time histogram for three similar prostate cases. Plans were created to irradiate the prostate and seminal vesicles with 2 Gy per fraction. a) MF of 2.445 and %LOT > pT-20 ms = 3.2%; b) MF of 2.007 and %LOT > pT-20 ms = 11.4%; c) MF of 1.809 and %LOT > pT-20 ms = 23.0%.

# CHAPTER 6

# A review on radiochromic film dosimetry for dose verification in high energy photon beams

Radiation Physics and Chemistry, 2021, volume 179, pages 109217

T Santos[a,b], T Ventura[b], MC Lopes[b]

[a] Physics Department, University of Coimbra, Coimbra, Portugal
[b] Medical Physics Department, IPOCFG, E.P.E., Coimbra, Portugal

## Abstract

A dosimetry system composed by Gafchromic™ EBT films and a flatbed document scanner is a valuable tool for patient-specific quality assurance (QA) of complex external beam radiotherapy treatment plans. However, it is still not widely used in clinical practice. The need for a comprehensive characterization of the film-scanner performance and the establishment of a complex dosimetry protocol may be one of the causes of this weak adhesion. The purpose of this work is to shed light on how to establish a reliable film dosimetry system for dose verification in external photon beam radiotherapy. The practical aspects to take into account are reviewed and the advantages and limitations of the different approaches are critically discussed according to the current state of the art, working as a guide for new users.

**Keywords:** Radiochromic film dosimetry, external photon beam radiotherapy, patient-specific QA

## 6.1 | Introduction

Radiochromic films are a versatile 2D dosimeter for many clinical tasks in radiation oncology given the associated characteristics, such as: an unrivaled high spatial resolution, near-water equivalent composition, weak energy dependence from kV to MV range, angular independence, and consistent response in an extended range of doses (Niroomand-Rad et al., 1998). Radiochromic films may also be immersed in water for short time periods, cut into several pieces and are relatively low cost.

Primarily used for external photon beam dosimetry (Distefano et al., 2017; Iqbal et al., 2018; Liu et al., 2015; Lopes et al., 2019; Santos et al., 2019a), radiochromic film based dosimetry systems are finding increasing application in brachytherapy (Ayoobian et al., 2016; Gholami et al., 2016), particle beam radiotherapy (Anderson et al., 2019; Avanzo et al., 2020; Kern et al., 2020), diagnostic radiology (Costa et al., 2017; Soliman et al., 2018), and radiobiological studies (Rutherford et al., 2019; Torres del Río et al., 2019). In external photon beam radiotherapy, radiochromic films commercially designated by Gafchromic™ EBT films and manufactured by Ashland, Inc., Wayne, NJ, USA have demonstrated to be a valuable tool for dose verification of intensity modulated radiotherapy (IMRT) and stereotactic treatment plans (Colodro et al., 2017; Cusumano et al., 2015; Marrazzo et al., 2015; Wen et al., 2016). IMRT and stereotactic treatment plans typically present steep dose gradients, use small fields and require an increased accuracy. Patient-specific quality assurance (QA) verifications should thus be performed prior to treatment to assess the agreement between the measured and the planned dose distributions (Hodapp, 2012). Despite the promising potential, radiochromic film based dosimetry systems are not a widely adopted tool for pre-treatment QA in clinics (Mehrens et al., 2020), probably due to what seems to be a cumbersome system characterization and the implementation of complicated dosimetry protocols. Indeed, to introduce a radiochromic film based dosimetry system in the clinical practice it is necessary to carefully select and characterize both the film and the scanner responses as well as to establish and follow a comprehensive and strict dosimetry protocol. The dosimetry protocol should give detailed instructions on films storage, handling, irradiation, scanning, image analysis, and absorbed dose determination (Devic et al., 2016). There are also many sources of uncertainty to consider for accurate dose measurements, arising from the film, film-scanner interaction and the scanner itself.

When Gafchromic™ EBT films are exposed to radiation, an image is formed through a post-irradiation process, resulting from a solid-state polymerization of the sensitive component (lithium-10,12-pentacosdiynoate – LiPCDA). The most widely used devices to measure the film response to radiation are flatbed color document scanners, mainly due to the low cost and ease of use. The measured response is commonly quantified in terms of the scanned pixel values (PV), optical density (OD), or the change of optical density of the film scanned before and after irradiation (netOD) (Aldelaijan and Devic, 2018). To relate the film response to absorbed dose, the first step is to establish a calibration curve per each scanner color channel (red – R, green – G and blue – B) through the irradiation of one or more films with known doses, under specific (reference) conditions, defined by dosimetry protocols such as TRS-398 (IAEA, 2001) or TG-51 (Almond et al., 1999). Once having a calibration curve, the response of any irradiated film from the same production lot can be converted into a dose map using either single channel or multichannel algorithms. Single channel methods are based on the application of the calibration curve corresponding to a given color channel. Multichannel algorithms combine the information provided by the different channels to determine a single dose value per pixel (Micke et al., 2011). Several multichannel methods have been proposed (Aldelaijan et al., 2019; Mayer et al., 2012; Méndez et

al., 2018, 2014; Micke et al., 2011; Pérez-Azorín et al., 2014) and demonstrated that they may contribute to an improved accuracy over single channel dosimetry, compensating for some sources of uncertainty such as film spatial heterogeneities (variations in sensitive layer thickness, scratches, fingerprints, dust), and scanner non-uniform response (Micke et al., 2011). Still, there are many factors that may vary between the calibration and the measurement phases. Those factors include for instance, post-irradiation darkening, changes in film response due to temperature, humidity and UV-light, intra-lot variabilities arising from differences in composition/thickness of the sensitive layer of the films and inter-scan variations of the scanner response (Aldelaijan et al., 2016; Lewis et al., 2012; Lewis and Devic, 2015; Palmer et al., 2015b; Rink et al., 2008). All these factors should be addressed when establishing a dosimetry protocol.

There are some software packages commercially available for film dosimetry, such as FilmQA PRO (Ashland ISP Inc., Covington, KY, USA), Radiochromic.com (a web application), or Verisoft (PTW-Freiburg, Germany) that may come with detailed instructions on how to establish an efficient protocol. Nevertheless, the users have an important role in their commissioning and implementation. If medical physicists teams prefer to take control on all the involved steps they should develop their own protocol. In any case, it is important that the users have a full insight and critical thoughts on every single step of the process.

Radiochromic film dosimetry has been a research area of interest, as demonstrated by the number of papers published on the topic over the past thirty years. The last review on film dosimetry was issued in early 2016, with only three bibliographic references after 2014 (Devic et al., 2016). The authors discussed the technical aspects associated with the implementation of a radiochromic film dosimetry system, presenting useful guidelines. From January 2015 to May 2020, dozens of peer-reviewed articles have been published on practical aspects of film dosimetry for high energy photon beams. This turns into a huge and complex amount of information that newcomers to the field should carefully read and analyse, especially when aiming to develop in-house protocols. With this concern, Howard et al. (2020) have recently published a methodology for film dosimetry strictly based on their own clinical experience.

The purpose of the present review is to shed light on how to establish or improve a reliable and accurate film dosimetry system for absolute dose verification in external photon beam radiotherapy. A literature search was conducted in PubMed and Scopus on peer-reviewed original research articles written in English published from January 2015 to May 2020. The used keywords were: “radiochromic” or “gafchromic” or “EBT” or “film dosimetry”. From the identified articles, only those referring factors related with the characterization and set-up of radiochromic film dosimetry systems for dose verification in high energy photons beams were considered. Fifty five articles were selected, forming the basis for this work. Yet, older publications since the issue of Gachromic™ EBT film are also referred either to provide a background on a certain subject or because the reported methods are still considered up to date.

The practical aspects related to film dosimetry, namely to the three main components, film model, scanner model and dosimetry protocol, are reviewed according to the current state of the art, highlighting the different issues to be considered and the available options, hopefully working as a guide for new users.

## 6.2 | Radiochromic film dosimetry system

### 6.2.1 | Radiochromic film models

The first radiochromic film models used for clinical dosimetry in radiotherapy (HD-180, MD-55, MD-55-2, HS) had several limitations, including a relatively low sensitivity, lack of uniformity,

small size and high price, preventing their generalized use (Devic, 2011). The release in 2004 of the EBT (External Beam Therapy) Gafchromic™ film (Ashland ISP Inc., Covington, KY, USA), with improved sensitivity and uniformity properties, represented a paradigm shift, definitely replacing the labor-intensive silver halide radiographic films in external beam radiotherapy dose measurements (Sankar et al., 2006).

#### 6.2.1.1 | Gafchromic™ EBT film

The EBT film was composed by two 17 μm thick active/sensitive layers, with a 6 μm surface layer in between, and coated with a 0.97 μm polyester sheet at each side. The active layers contained the radiation-sensitive component (lithium-10,12-pentacosdiynoate in crystalline form), dispersed on it (Devic et al., 2005). Since its release, the properties of EBT film were extensively studied, seeking continuous improvements, which led to the development of EBT2, EBT3 and EBT-XD films in 2009, 2011 and 2015, respectively.

#### 6.2.1.2 | Gafchromic™ EBT2 film

The EBT films were replaced by the model Gafchromic™ EBT2 in which the active part of the films was reduced to a single layer 28 μm thick and a thin topcoat applied to a 175 μm polyester substrate. Moreover, the coated layers are over-laminated with polyester. According to manufacturer, some of its main advantages include a higher tolerance to visible/ultraviolet light exposure and it can correct non-uniformity of the active layer thickness using a yellow marker dye. For these films, the non-symmetric structure leads to a new dependence on the 'front-back' film orientation (Carrasco et al., 2013).

#### 6.2.1.3 | Gafchromic™ EBT3 film

The EBT3 films can be considered a refinement of the EBT2 model with a symmetric configuration, consisting of the same active layer as ETB2 films but sandwiched between two polyester layers (125 μm) – Figure 6.1 – that promote the stiffness of the film and permit its use in water. The active layer is made of the active compound (LiPCDA crystals with 1-2 μm in diameter and 15-25 μm in length (Lewis and Chan, 2016)), the marker dye, and other stabilizers and additives that minimize the energy dependence of the film (Ashland, 2020a). The polyester substrate has a special surface treatment containing microscopic silica particles, that prevents the formation of Newton ring-like artifact during scanning (Casanova Borca et al., 2013), reported for previous EBT models (Ashland, 2020a). EBT3 symmetric structure contributes to reduce the face-up/down scanning response variation (Casanova Borca et al., 2013; Moylan et al., 2013). Gafchromic™ EBT3 film optimal dose range varies from 0.2 to 10 Gy, but using the information provided by the three color channels it is possible to extend the film dynamic range up to 20 Gy (Ashland, 2020a). This extension comes along with an increased uncertainty at high doses (> 10 Gy) mainly due to the shallow slope of the calibration curves (Ashland, 2020a; Palmer et al., 2015b).

#### 6.2.1.4 | Gafchromic™ EBT-XD film

The EBT-XD film, released in 2015, presents an optimal performance in the dose range 0.4-40 Gy, with declared better performance for hypofractionated treatments verification, including stereotactic radiosurgery and stereotactic body radiotherapy. The EBT-XD has the same structure as EBT3 film (Ashland, 2020b; Palmer et al., 2015b), with a thinner active layer (25 μm) – Figure 6.1 – and different chemical composition: the LiPCDA crystals are 1-2 μm in diameter and 2-4 μm in length (Lewis and Chan, 2016) – Table 6.1. This reduction in size of the active particles decreases scattering and light polarization (León-Marroquín et al., 2019a).

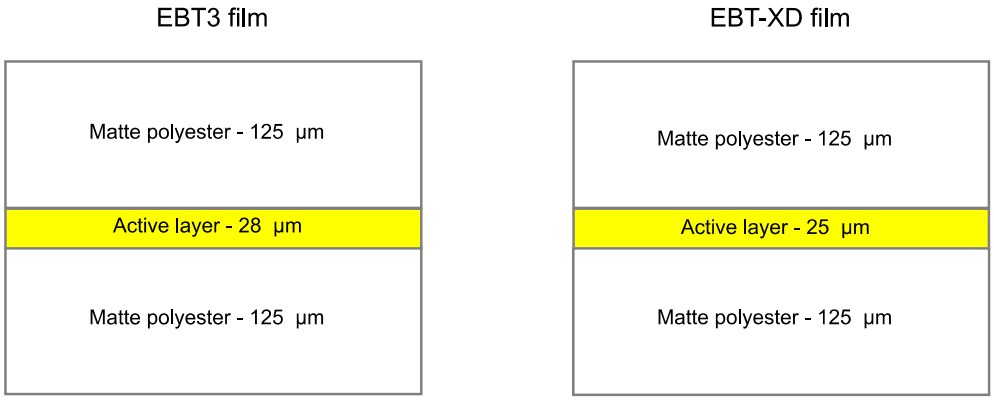


**Figure 6.1** – Structure of Gafchromic™ EBT3 and EBT-XD films. Adapted from (Ashland, 2020a, 2020b).

**Table 6.1** – Dose range (optimum and dynamic), and active layer composition of Gafchromic™ EBT3 and EBT-XD films. Adapted from (Palmer et al., 2015b).

| | | | **Active layer composition (atomic %)** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **Film model** | **Optimum dose range** | **Dynamic dose range** | H | Li | C | N | O | S | Na | Cl | Al |
| **EBT3** | 0.1 to 10 Gy | 0.1 to 20 Gy | 56.8 | 0.6 | 27.6 | - | 13.3 | - | - | - | 1.6 |
| **EBT-XD** | 0.4 to 40 Gy | 0.1 to 60 Gy | 57.0 | 0.6 | 28.5 | 0.4 | 11.7 | 0.1 | 0.1 | 0.1 | 1.4 |

#### 6.2.1.5 | Comparison between EBT3 and EBT-XD films

Several comparative studies have recently been published on EBT3 and EBT-XD film properties characterization for high energy photon beams (Grams et al., 2015; Khachonkham et al., 2018; León-Marroquín et al., 2019a; Miura et al., 2016; Palmer et al., 2015b; Schoenfeld et al., 2016a). Aspects such as the absorption spectra, dose response, sensitivity, and energy and dose rate dependence were analysed. León-Marroquín et al. (2019a) in a comparison of the spectroscopic characteristics of EBT3 and EBT-XD films, have shown that the absorption spectra of both films present two main peaks centered at around 636 nm and 585 nm. These absorption peaks are indeed characteristic of the EBT model. The spectroscopic characteristics of the films do not change with beam energy and dose rate. However, they seem to be dependent on dose level and films production lot (León-Marroquín et al., 2019a, 2019b). Palmer et al. (2015b) have evaluated the calibration/dose-response curves for EBT3 and EBT-XD films in a dose range up to 40 Gy. They demonstrated that EBT-XD films are less dense for the same dose than EBT3, as a higher dose is necessary to reach the same response. Moreover, the red and green channels dose-response curves for EBT-XD are well separated from each other and have a steep gradient for high doses, in contrast to EBT3 curves which present a shallow slope, tending to saturation. Both curves separation and their gradient are crucial for improved accuracy in single- and multi- channel dosimetry techniques (Palmer et al., 2015b). To illustrate this, an example of R, G, B dose-response curves obtained for both EBT3 and EBT-XD films is provided in Figure 6.2.

Palmer et al. (2015b) have also analysed the gradient of the dose-response curves for each color channel to establish the dose regions of maximum sensitivity for each film model. They reported that at doses higher than approximately 5 Gy and 8 Gy, the gradient of the EBT-XD dose-response curves corresponding to the red and green channels, respectively, surpass that of EBT3 films. The higher slopes presented by EBT-XD films in the red and green channels for the high dose range, make them more reliable for high doses verification (Ashland, 2020b). For EBT3 films, the green channel has shown an improved sensitivity for doses higher than 8-10 Gy (Khachonkham et al.,

2018; León-Marroquín et al., 2019a; Wen et al., 2016). Regardless the film type, the blue channel dose-response curve exhibits a lower gradient at any dose. Indeed, its response is more dependent on the thickness of the active layer rather than on the dose as the yellow marker dye absorbs light in the blue part of the spectrum. These findings were corroborated by other authors (Khachonkham et al., 2018; León-Marroquín et al., 2019a; Lewis and Chan, 2015; Miura et al., 2016), evidencing the higher sensitivity of EBT3 films for low doses (< 10 Gy) and the improved sensitivity of EBT-XD films for high doses (> 10 Gy).

The response of both EBT-XD and EBT3 films is weakly dependent on energy and dose rate (DR) in a range from 100 kV to MV photon energies. Miura et al. (2016) have analysed the EBT-XD film response dependence on energy and dose rate by irradiating film strips at doses from 0.5 to 40 Gy for 6 MV (DR=600 MU/min), 6 MV FFF (DR=1400 MU/min), 10 MV (DR=600 MU/min) and 10 MV FFF (DR=2400 MU/min) photon beams. The netOD differences were within 1.5%. Grams et al. (2015) reported differences in OD below 0.8% for EBT-XD film strips irradiated with 6-18 MV photon beams with doses up to 30 Gy. Khachonkham et al. (2018), in a comparative study, have shown that both EBT3 and EBT-XD films exhibit a similar response in the kV to MV photon energies range.

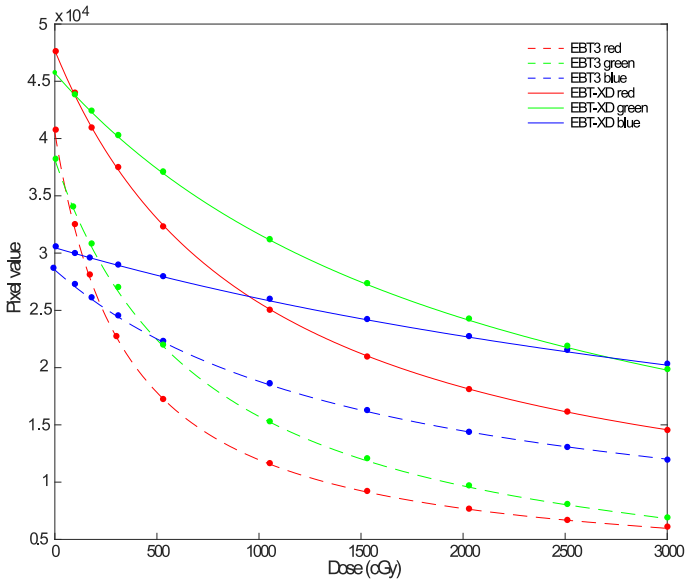


**Figure 6.2** – Red, green and blue dose-response curves for Gafchromic™ EBT3 and EBT-XD films.

### 6.2.2 | Flatbed document scanners

The most widely used devices to measure the film response to radiation are the commercially available flatbed document scanners, adapted to operate in both reflection and transmission modes. Transmission mode is the most commonly employed scanning method in clinical practice. Nevertheless, some authors suggested that using reflection mode could be advantageous due to the improved film sensitivity and lower cost of the scanners without the transparency option (Méndez et al., 2013; Papaconstadopoulos et al., 2014; Park et al., 2012). Aldelaijan and Devic, 2018 in a recent comparison between the response of EBT3 film for doses up to 11 Gy in reflection and transmission modes, confirmed that the reflection mode provides an improved sensitivity and accuracy, but in a limited dose range (up to 5 Gy), while the associated uncertainties tend to be higher (Aldelaijan and Devic, 2018). In the present review, we will just focus on transmission mode.

The scanners used in film dosimetry are typically equipped with a linear charge couple device (CCD) array detector, and depending on the scanner model, the light source may be a fluorescent

lamp or light emission diodes (LED). Regardless the light source technology, the measured signal results from a convolution of three factors: the light source emission spectrum, the film absorption spectrum and the sensitivity of the linear CCD detector array (Aldelaijan et al., 2019). This means that, even for the same scanner model, the response and sensitivity of the film-scanner system may be different. Flatbed scanners with a fluorescent lamp such as Epson Expression 10000 XL/11000 XL and Epson Perfection V700 have been utilized in film dosimetry for many years. LED-based scanner models, like Epson Expression 12000 XL and Epson Perfection V800 were recently released, replacing the former ones in the market. The performance of both scanner types seems to be generally comparable, yet the LED-based ones appear to present a more pronounced non-uniform response (Lárraga-Gutiérrez et al., 2018; Shameem et al., 2020), but the associated noise tends to be lower (González-López et al., 2017; Vera-Sánchez et al., 2018). Shameem et al. (2020), for instance, showed that the Epson Perfection V800 scanner presents a non-uniform response in the three color channels about 2% greater than the V700 scanner. González-López et al. (2017), found the random noise in the Epson Perfection V800 to be about one half of that of the Epson Expression 10000 XL, for all color channels and resolutions under study.

The use of flatbed scanners in clinical practice, demands a comprehensive film-scanner performance characterization, and the establishment of a scanning protocol, including optimal parameters like color bit depth and spatial resolution (Tagiling et al., 2018). Aspects such as scanner lamp warm-up, noise, inter-scan variations, orientation effect and scanner non-uniform response should also be addressed (Devic et al., 2016; Ferreira et al., 2009; Paelinck et al., 2006).

It is common practice to use scanners in 48-bit RGB mode with 16 bits per color channel, with all color corrections turned off. Regarding the scanning spatial resolution however, a consensus does not exist, with values as different as 50 and 254 dpi being reported in the current literature (Akdeniz and Yegingil, 2019; Aldelaijan et al., 2016; Bennie and Metcalfe, 2016; González-López et al., 2017; León-Marroquín et al., 2019a; Méndez et al., 2018, 2016; Palmer et al., 2015b; Ruiz-Morales et al., 2017; Tagilingl et al., 2018). Anyway, 72 dpi seems to be the most widely adopted scanning resolution. On one hand, the selection of the scanning resolution determines the spatial resolution of the measured dose distribution. But on the other hand, it affects the noise level in the scanning image, i.e., the higher the resolution the higher the noise variance (González-López et al., 2017; Méndez et al., 2016; Vera-Sánchez et al., 2016). It is also well described that the noise level may eventually impact the agreement between the measured and planned doses, during the pre-treatment verification phase (Huang et al., 2014). Therefore, the use of de-noising techniques such as taking the average image of multiple scans – time averaging – and/or the application of a 2D adaptive Wiener filter over the scanned images has been highly advised (Devic et al., 2016; Ferreira et al., 2009; Vera-Sánchez et al., 2016). There are authors that use median filters instead (Méndez, 2015; San Miguel et al., 2018). Moreover, scanning at a high resolution (> 100 dpi) and then downscaling to the deemed resolution can be an advantageous option, as suggested by Devic et al. (2016) and Méndez et al. (2016). This approach allows to characterize each image pixel by a mean value and standard deviation, appraising noise and spikes in the measured signal (Devic et al., 2016). In addition, it seems to prevent the formation of the so-called “grid pattern” artifact that results from the periodical variation of the noise in both scanner bed axis and has been identified for scanning resolutions up to 96 dpi (Méndez et al., 2016). The dosimetric impact of the “grid pattern” was studied by Méndez et al. (2016) and has shown to depend on the color channel, dose, scanner repeatability and resolution. The blue channel is the most affected, with the reported relative dose uncertainty varying between 6% and 8%, for a dose of 100 cGy and a scanning resolution of 96 dpi. For the red and green channels, the relative dose uncertainty was mostly below 2%, for all doses (100 cGy, 200cGy, 300 cGy and 400 cGy) and scanning resolutions under study (50 dpi, 72 dpi, 96 dpi and 150 dpi). Some investigators reported on the scanner warming up effect

due to temperature changes of the scanner lamp. To ensure a stabilized response, it is recommended to perform a certain number of pre-scans prior to its use, as soon as the in-built scanner warm-up time has elapsed, or after long pauses (Ferreira et al., 2009). The number of required pre-scans should be assessed for each scanner by the users.

Inter-scan variations of the scanner response have recently been addressed (Lewis and Devic, 2015; Méndez et al., 2016; Palmer et al., 2015a; Vera-Sánchez et al., 2016) and various methods have been proposed to compensate for them. Inter-scan variations may be caused by changes in the stability of the CCD detectors and of the electronic circuitry of the scanner, effect of temperature differences, and variations in the film-to-light source distance (Lewis and Devic, 2015). Such changes may compromise the dose-response relationship between the calibration and the subsequent measurement scans and thus should be corrected for. To mitigate the effect of film-to-light distance variation that is mainly due to the effect of the natural film curling, Palmer et al. (2015a) and Lewis and Devic (2015) advised the use of a glass compression plate (3-4 mm thick) on top of the films to be scanned. Regarding the other sources of variation, corrections may be carried out by means of a control film piece or a recalibration process, as discussed in section 6.2.3.1.

Two well-known artifacts associated with film digitization in flatbed scanners are the so-called 'orientation effect' and 'lateral response artifact'. The orientation effect is characterized by a change in response as a function of the orientation of the film on the scanner bed, whose magnitude depends on the dose and the scanner (Schoenfeld et al., 2016b, 2014; van Battum et al., 2016). A difference in PV of approximately 8% for a rotation of 90° has been reported by Khachonkham et al. (2018) for EBT3 films at a dose of 2000 cGy, scanned using an Epson 11000 XL. The orientation of the film during exposure has not discernible impact on the dose response. Flatbed scanners present a non-uniform response orthogonally to the scanning direction, which induces the so-called lateral response artifact (LRA). The LRA is characterized by a parabolic decrease in PV as a function of the lateral distance from the center of the scanning area (Schoenfeld et al., 2016b; van Battum et al., 2016). The response difference at a given lateral position increases with decreasing PV, i.e. increasing dose, being very pronounced for high doses (above 10 Gy). Palmer et al. (2015) verified a reduction in the PV of about 28% for EBT3 films irradiated to 20 Gy and placed at the edge (15 cm off-axis) of an Epson 11000 XL scanner. Lewis and Chan (2016) have reported that the differences in response are more noticeable for the red than for the green and blue channels, and demonstrated that the LRA effect can be mitigated using multichannel dosimetry. This effect varies from scanner to scanner but typically scanners with a large scanning area are preferred, as the LRA effect is reduced when films are digitized at the center of the bed. Multiple scanner non-uniform response correction methods can be found in the literature (Ferreira et al., 2009; Lewis and Chan, 2015, 2016; Li et al., 2017; Méndez, 2015; Menegotti et al., 2008; Pérez-Azorín et al., 2016; Poppinga et al., 2014; Wen et al., 2016). Wen et al. (2016), for instance, generated a non-uniform response correction map for the scanner based on the registration and comparison of the film measured dose with the dose map measured by an ionization chamber array, taken as reference. Lewis and Chan (2015) have proposed a simple correction method that can be established at the moment of the scanner commissioning, as it demonstrated to be independent of the film production lot and dose. This method assumes that for each color channel X, the measured response of a film irradiated to a known dose D and at a lateral position L is related to the response at the scanner center C by: $Response(C, D, X) = A_{L,X} + B_{L,X} \cdot Response(L, D, X)$, where $A_{L,X}$ and $B_{L,X}$ .are correction coefficients. The correction coefficients are determined from the measured response at the scanner center and at least six lateral positions of two films exposed to known doses, such as 0 Gy and the maximum dose of interest for clinical use. The values for intermediate lateral positions are determined by linear interpolation (Lewis and Chan, 2015).

The orientation and LRA effects also depend on the film model, as a significant reduction for EBT-XD films in comparison with EBT3 films has been noticed (Khachonkham et al., 2018; Lewis and Chan, 2016; Palmer et al., 2015b; Schoenfeld et al., 2016b; van Battum et al., 2016). Khachonkham et al. (2018) reported that for an 8 cm lateral off-axis position in an Epson 11000 XL scanner, the difference in PV relative to the center for a dose of 200 cGy was 5% for EBT3 and 2% for EBT-XD films. For a dose of 2000 cGy, the difference was 17% for EBT3 and 6% for EBT-XD. Regarding the orientation effect measured in PV, a reduction of 3.1% for EBT-XD in comparison to EBT3 was verified for a dose of 2000 cGy. Such differences can be justified by the decrease of the polarization effects and practically complete loss of the anisotropy of the light scattered by an EBT-XD exposed film, as pointed by Schoenfeld et al. (2016a).

The orientation effect can be mitigated by adopting a constant film scanning orientation, whereas the need for a LRA correction has to be carefully assessed, based on the dimensions of the irradiated area, dose of interest, film type (EBT3 or EBT-XD) and applied dosimetry method (either single channel or multichannel) (Li et al., 2017; Micke et al., 2011; Palmer et al., 2015b).

### 6.2.3 | Dosimetry protocol

A film dosimetry protocol consists in a set of instructions on film storage, handling, irradiation, scanning, image analysis, absorbed dose determination, and uncertainty estimation and should be strictly followed during both calibration and measurement phases (Niroomand-Rad et al., 1998). Films should be handled with gloves, kept in black envelopes when not being used to minimize the exposure to light and may be stored at room ambient temperature (20°-25°C), or ideally at refrigerator temperature (Ashland, 2020c).

#### 6.2.3.1 | Calibration phase

Gafchromic™ EBT films used in the clinical routine are typically available in boxes of 25 sheets (8×10 inches in size) from a certain production lot. As the chemical composition of the films and the thickness of the active layer may vary between lots, to convert the measured film response to absorbed dose, a dose-response curve per lot for each channel should be established. That is typically accomplished through the irradiation of strips from one or more films with known doses, under specific (reference) conditions, defined in dosimetry codes of practice like TRS-398 (IAEA, 2001) or TG-51 (Almond et al., 1999). The calibration process should ideally be repeated for each new box, even when the lot number is the same (Devic et al., 2016). Plan-based calibration methods consisting in the irradiation of a film sheet with different dose levels have also been used and demonstrated to be faster and more convenient than the strip-based method though less precise, being not generally so common (Li et al., 2017; Méndez et al., 2013). In the present review, we will just focus on strip-based calibration.

The definition of dose-response functions specific to each film lot/box, scanner, scanning conditions and color channel, is a critical step in film dosimetry. The users have to: 1) decide how many strips to irradiate and to what doses; 2) choose a quantity to express the film response to radiation; and 3) find an analytical function for the calibration curve fitting. Regarding the number of strips, there is not a rule, so far, but they should ideally fit in a row in a single scan, to avoid inter-scan variations. Either a guillotine paper cut or a regular scissors with sharp edges can be used to cut the film in strips. When using a scissors, extra care should be taken as delamination of the film may cause unusable edge effects, making the measurements close to the film edges unreliable (Devic et al., 2016).

Prior to the calibration strips irradiation, the radiation beam output must be measured. During the calibration strips irradiation, one option is to place an ionization chamber below the film plane

to determine the dose delivered to the film, following a dosimetry reference code of practice. Calibration film strips are then digitized, according to a previously established dosimetry protocol, taking into account the considerations made in Section 6.2.2. Briefly, the users have to select the color bit depth, spatial resolution, and turn-off all color corrections options. Films should be centered at the scanner bed, forming a single row in the scanning direction, and at a consistent orientation. A glass plate should be placed on the film top for flattening its surface (Palmer et al., 2015a).

The elapsed time between irradiation and scanning is an important factor to consider in film dosimetry, due to the film post-irradiation darkening. Studies on the characterization of the post-irradiation darkening of EBT3 and EBT-XD films, from few minutes up to 72 h, have shown that the response changes rapidly in the first 6 h after exposure, while it nearly stabilizes after 24 h, with differences in netOD from 24 h to 48 h of about 2% (Khachonkham et al., 2018; Miura et al., 2016). In order to guarantee an adequate stability for film dosimetry, many authors recommend to scan films 24 h or even 48 h after irradiation. Moreover, this waiting time should be ideally the same for both calibration and subsequent irradiated films in order to further minimize dose uncertainties. In alternative, a recalibration protocol can be used. Lewis et al. (2012) established a measurement protocol – "Efficient protocol" – based on the scan of the irradiated film (measurement film) together with two reference strips (preferably cut from the same film), one irradiated to a known dose and the other non-irradiated that are used to adjust the established dose-response curves. The authors demonstrated that this protocol allows to obtain accurate dose values (dose error < 0.5%), after a period equal to four times the waiting time between the irradiation of the measurement film and the reference strips. Similar findings were reported by Shimohigashi et al. (2015) and Calvo-Ortega et al. (2017) in a comparison of the accuracy achieved using the Efficient protocol and the conventional approach of scanning the films 24 h after irradiation, in dose verification of test and clinical IMRT plans (Calvo-Ortega et al., 2017; Shimohigashi et al., 2015). This is of great benefit when performing patient-specific QA verifications, as the processing time may be reduced from days to a few hours.

After scanning, the images should be processed using an imaging processing software like MATLAB (Mathworks, Natick, MA, USA) dedicated routines, or directly in the software packages commercially available for film dosimetry. The amount of noise should be reduced, by taking the mean of several scans (typically 4 or 5) and/or applying a 2D adaptive Wiener filter (Devic et al., 2016; Ferreira et al., 2009; Vera-Sánchez et al., 2016). A region of interest (ROI) is defined at the center of the exposed areas, and the average PV recorded. The ROI dimensions vary widely across the published protocols, with sizes as diverse as $0.5\times0.5$ cm$^2$ and $4\times4$ cm$^2$ (Akdeniz and Yegingil, 2019; Howard et al., 2020; Méndez et al., 2018; Palmer et al., 2015b; Ruiz-Morales et al., 2020; Spelleken et al., 2018; Wen et al., 2016). That choice is rather subjective and depends on the dimensions of the irradiated strips, the field size, but also on the acceptable dose uncertainty (Bouchard et al., 2009). According to the EBT films manufacturer's (Micke et al., 2011), the selected area should be such that the measured values are not particularly affected by any disturbance, however, the beam homogeneity should also be taken into account. Devic et al. (2016) proposed an alternative approach based on the selection of multiple ROIs around the center of the exposed area. The final response is calculated as a weighted mean over those ROIs values.

The film response to radiation has been expressed mainly in terms of PV, OD and netOD, though some authors consider other quantities such as transmittance (Akdeniz and Yegingil, 2019; Devic et al., 2016; Méndez, 2015; Méndez et al., 2018; Micke et al., 2011). The use of netOD has demonstrated to partially mitigate spatial heterogeneities, mainly differences in the film active layer thickness, when there is a perfect match between the images of the film scanned before and after irradiation (Méndez, 2015). Méndez et al. (2018) reported a generally increased accuracy when

using netOD in comparison with OD and PV, for both single and multichannel methods. However, this choice has associated a greater uncertainty and implies an increased workload, which may not be justified specially when using triple channel dosimetry techniques (as discussed later) (Méndez, 2015; Méndez et al., 2018). Aldelaijan et al. (2018) have explored the option of using the readout from a non-irradiated strip scanned together with the irradiated films, for background signal definition (fog). They found negligible differences in terms of accuracy and uncertainty between both approaches. The authors also reported that the use of normalized PV calculated as:

$$\overline{PV} = \frac{PV^{unirr}}{PV^{irr}} - 1 \tag{6.1}$$

linearizes the dose-response and contributes to an increased sensitivity over netOD based dose-response functions (Aldelaijan and Devic, 2018). Linearization of the calibration functions allows to perform relative dose measurements using film with no need to go through the entire calibration process (Devic et al., 2019, 2012).

Various analytical functions for the calibration curve fitting of the measured dose-response data points have recently been applied, including multi-parameter polynomial, power, exponential, logarithmic, and rational forms, depending on the used response quantity and on the independent variable – Table 6.2. The most popular are rational functions that present a qualitative behavior similar to that of the film response to radiation, which may contribute to reduce the number of calibration points (Lewis et al., 2012). Regardless the functional form, the most suitable calibration function is the one that monotonically varies with increasing dose and minimizes the relative fitting uncertainty, as stated by Devic et al. (2016). New functional forms of calibration curves have also been recently published (Aldelaijan et al., 2019; Campajola et al., 2017; Casolaro et al., 2019; López-Fidalgo and Amo-Salas, 2020; Martin-Viera Cueto et al., 2015; Spelleken et al., 2018; Tamponi et al., 2016). Aldelaijan et al. (2019) proposed a single R-G-B dose-response function that is based on a weighting sum of the normalized PV obtained for the three channels, and linearizes the response for doses up to 40 Gy. Méndez et al. (2018) and Wen et al. (2016) use natural cubic splines to model the dose-response functions per channel. The dose accuracy and precision achieved using some of these calibration curves have been recently evaluated (Akdeniz and Yegingil, 2019; Aldelaijan and Devic, 2018; San Miguel et al., 2018; Tamponi et al., 2016).

#### 6.2.3.2 | Measurement phase

Once having a calibration curve per color channel, they can be used for the determination of an unknown dose distribution. The measurement films should be digitized using the same scanning protocol as for the calibration strips. The image analysis should also be performed in the very same way. To convert the film response into dose, one of two formalisms have been employed so far, which are single channel and multichannel methods. Multichannel dosimetry methods have shown potential to improve the dosimetric accuracy over single channel dosimetry, compensating for variations in the thickness of the film active layer, non-uniform response of the scanner and other artifacts such as fingerprints, scratches and dust (Akdeniz and Yegingil, 2019; Dufek et al., 2019; Marrazzo et al., 2015; Mayer et al., 2012; Méndez, 2015; Méndez et al., 2014; Micke et al., 2011; Pérez-Azorín et al., 2016; San Miguel et al., 2018). Yet, there are other factors, such as the post-exposure growth and film-scanner response, which may change between the calibration and the measurement phases that should be accounted for to obtain accurate dose values. The mitigation of these effects may be accomplished by means of two different approaches as the use of a control film piece or recalibration methods (Aldelaijan et al., 2016; Lewis et al., 2012; Lewis and Devic, 2015; Méndez et al., 2016; Ruiz-Morales et al., 2017).

**Table 6.2** – Review of calibration functions used in radiochromic film dosimetry, where x is the independent variable, y is the dependent variable, and a, b, c, d are fitting parameters.

| Function | | Response | Independent variable | References |
|---|---|---|---|---|
| **Rational** | $y = \frac{ax}{bx+1}$ | net transmittance | response | (Aldelaijan et al., 2016; Devic et al., 2016) |
| | | net transmittance, netOD, normalized PV | response | (Aldelaijan and Devic, 2018) |
| | $y = \frac{a+bx}{x+c}$ | PV | dose | (Bennie and Metcalfe, 2016) |
| | | OD | dose | (Shimohigashi et al., 2015) |
| | | netOD | response | (San Miguel et al., 2018) |
| | $y = a + \frac{b}{x-c}$ | PV | dose | (Lewis and Devic, 2015; Marrazzo et al., 2015; Palmer et al., 2015b; Ruiz-Morales et al., 2020, 2017; Vera-Sánchez et al., 2018, 2016) |
| | $y = \frac{ax}{x-b}$ | netOD | response | (San Miguel et al., 2018) |
| **Power** | $y = ax + bx^n$ | netOD | response | (Devic et al., 2016; Lárraga-Gutiérrez et al., 2018; Ruiz-Morales et al., 2020, 2017; Tagiling et al., 2018) |
| | | | dose | (León-Marroquin et al., 2016; León-Marroquín et al., 2019a, 2018) |
| **Polynomial** | $y = \sum_{n=1}^{3} b_n x^n$ | OD | response | (Chung et al., 2016; Ruiz-Morales et al., 2020, 2017) |
| | | netOD | response | (San Miguel et al., 2018) |
| | $y = \sum_{n=0}^{4} b_n x^n$ | PV | response | (Méndez, 2015) |
| | | netOD | response | (San Miguel et al., 2018) |
| | $y = \sum_{n=1}^{5} b_n x^n$ | OD | response | (Campajola et al., 2017) |
| **Logarithmic** | $\frac{\ln(\frac{b+x}{b})}{a}$ | netOD, net transmittance | dose | (Akdeniz and Yegingil, 2019) |
| **Exponential** | $y = a \cdot e^{bx} + c \cdot e^{dx}$ | PV | dose | (Khachonkham et al., 2018) |
| | $y = c - a \cdot e^{-bx}$ | OD | dose | (Li et al., 2017) |

**Single and multichannel dosimetry methods**

In single channel methods, the dose value at each pixel in the scanned images is obtained by the direct application of the dose-response curve corresponding to the chosen channel (R, G, B). The suitable channel depends on the dose range of interest and on the film model (either EBT3 or EBT-XD), as discussed in Section 6.2.1. Multichannel methods consider data from the different channels to derive a single dose value, aiming to compensate for spatial inhomogeneities of the film-scanner response, particularly differences in the film active layer thickness and lateral response artifact (Mayer et al., 2012; Méndez, 2015). Whereas triple channel methods combining the information of the three color channels are the most used, green to blue, red to blue ratios have also

been utilized (García-Garduño et al., 2016; Mayer et al., 2012; Méndez et al., 2014, 2013). The simplest proposed triple channel method consists in taking the average or weighted average of the dose derived from all channels (Mayer et al., 2012; Méndez et al., 2013). Nonetheless, the current standard are perturbation models, which turn to be conceptually and/or mathematically more complicated, preventing their generalized implementation and full understanding.

Perturbation models assume that the estimated dose per pixel (for each channel) deviates from the "true" dose value due to a small local perturbation of the film-scanner response (Méndez, 2015). Micke et al. (2011) were pioneers in this field, defining a model that demonstrated to be effective in separating the non-dose dependent artifacts due to the pair film-scanner from the dose map. This method is implemented in FilmQA Pro software (Ashland, 2020d), made available by the film manufacturer, and is worldwide used in the clinical practice (Dufek et al., 2019; Howard et al., 2020; Iqbal et al., 2018; Khan et al., 2019; Marrazzo et al., 2015). The model definition presented by Micke et al. (2011) was rather conceptual which lead Mayer et al. (2012) to publish a closed-form solution that can easily be programmed. Méndez et al. (2014) moved one step forward defining a generalized channel-independent perturbation (CHIP) model that included the Micke-Mayer model as a particular case. Basically, assuming that the perturbation is small and approximately equal for all channels, the absorbed dose D at each film pixel $r$ can be derived from:

$$\left\{ \begin{array}{l} D(r) = D_R\big(z_R(r)\big) + \dot{D}_R(z_R))\Delta(r) + \epsilon_R(r) \\ D(r) = D_G\big(z_G(r)\big) + \dot{D}_G(z_G))\Delta(r) + \epsilon_G(r) \\ D(r) = D_B\big(z_B(r)\big) + \dot{D}_B(z_B))\Delta(r) + \epsilon_B(r) \end{array} \right\} \tag{6.2}$$

where $D_k$ is the calibration function for each channel ($k = R, G, B$), $z_k$ is the measured film response, $\dot{D}_k$ is the first derivate of $D_k$ with respect to $z_k$, $\Delta(r)$ represents the perturbation and $\varepsilon_k$ is an error term accounting for the difference between the "true" dose value, $D(r)$, and the dose obtained after correcting for the perturbation (Méndez et al., 2014). As $\Delta(r)$ and $\varepsilon_k(r)$ are unknown, the determined dose values $D(r)$ depend on the assumed probability density functions (pdf) for both terms and the authors propose and analyse some examples. The Truncated Normal distribution model assumes that the pdf of the perturbation term $\Delta$ follows a truncated normal distribution, and the $\varepsilon_k$ may be different. In the Micke-Mayer model the pdf of $\Delta$ is a uniform distribution and the pdf of all $\varepsilon_k$ are equal (Méndez et al., 2014). The general model presented here can be classified as an additive channel-independent perturbation model, as the perturbation in the response is additive and approximately equal for all the channels (Méndez et al., 2018).

Pérez-Azorín et al. (2014) have also defined an additive perturbation model, but the perturbation is considered to be a sum of two factors, one that depends on the color channel and another that is channel-independent. Whereas the authors have claimed a potential improved accuracy over the Micke-Mayer method (Mayer et al., 2012; Micke et al., 2011), it has not been widely adopted.

Méndez et al. (2018) have recently defined a new category of channel-independent perturbation model – the multiplicative CHIP perturbation model – which considers that the perturbation is multiplicative and approximately equal for all the channels. Once again, the determined absorbed dose depends on the assumed pdf for the perturbation and error terms. The same authors also proposed a novel method called "Multigaussian", which builds on the assumption that the pdf of the measured response for all channels obeys a multivariate Gaussian distribution (Méndez et al., 2018; Vera-Sánchez et al., 2018). The Multigaussian method is implemented in Radiochromic.com, a cloud computing web application (Radiochromic.com, 2020).

Multiple studies have been conducted on the comparison of the achieved accuracy when using both single and triple channel dosimetry methods (Akdeniz and Yegingil, 2019; Dufek et al., 2019; Marrazzo et al., 2015; Mayer et al., 2012; Méndez, 2015; Méndez et al., 2018, 2014; Micke et al., 2011; Pérez-Azorín et al., 2014; Ruiz-Morales et al., 2020; San Miguel et al., 2018). San Miguel et al. (2018), for instance, have compared the red channel dosimetry results against four perturbation models: the Micke-Mayer's, (Mayer et al., 2012; Micke et al., 2011), the Truncated Normal distribution (Méndez et al., 2014), the Pérez-Azorín's model (Pérez-Azorín et al., 2014) and a hybrid solution combining both Méndez's and Pérez-Azorín's ones. The authors analysed the gamma passing rates obtained for IMRT and volumetric modulated arc therapy (VMAT) plans from multiple sites, including conventional and hypofractionated regimens, using EBT3 films. They found that triple channel methods contributed to a better agreement between measured and calculated dose distributions, with the Pérez-Azorín model providing the most consistent results (San Miguel et al., 2018). Méndez et al. (2018) compared the accuracy achieved with the Multigaussian method against additive and multiplicative CHIP models as well as single channel dosimetry, for doses up to 16 Gy (with EBT3 films) and 32 Gy (with EBT-XD films), by considering the relative difference between film doses and doses measured with MatriXX (taken as reference). Globally, the Multigaussian model provided the most accurate dose estimates. Mean absolute errors as low as 1% and 0.8% were reported for the Multigaussian method using PV and net response, after applying lateral corrections (Méndez et al., 2018). The application of lateral non-uniformity corrections, which depend on the scanner width and film model being analysed, contributed to obtain more accurate film doses for all methods under study. The improvement was especially pronounced for single channel methods (R, G, B), with a reduction in the mean absolute error of about 2% (averaged over all channels). Using the response change instead of the raw PV improved the results for single channel methods, being the mean absolute error reduction of about 1%, on average. However, this was not verified for all triple channel algorithms under study (Méndez et al., 2018).

A comprehensive evaluation of the ability of multichannel methods to reduce the effect of different sources of noise in comparison with single channel dosimetry was performed by González-López, et al. (2017). The authors concluded that if de-noising techniques are not applied, random noise may be a significant source of uncertainty in radiochromic film dosimetry even when using multichannel dosimetry (González-López et al., 2017).

**Use of a lot control film piece and recalibration methods**

Lewis and Devic (2015), Aldelaijan et al. (2016) and Méndez et al. (2016) recommended the use of an unexposed film piece from a given box/lot, scanned at a fixed position to correct for eventual film-scanner response changes between the calibration and the subsequent scans of films from that same box/lot. Using such control film piece, the change of the response (per color channel) between scans may be used to derive multiplicative factors to correct the PV of the scanned images (Aldelaijan et al., 2016; Méndez et al., 2016). However, the application of these correction factors should be based on an accuracy vs uncertainty analysis (Aldelaijan et al., 2016).

The first recalibration method to be proposed was the so-called "Efficient protocol" (Lewis et al., 2012). According to this protocol, a generic calibration curve is established per lot/box for each color channel by fitting a rational function to the dose-response data:

$$\mathrm{N}(D) = \frac{\mathrm{PV}(D)}{\mathrm{PV}(0)} = a + \frac{b}{D + c} \tag{6.3.1}$$

$$\mathrm{N}(D) = \frac{\mathrm{PV}(D)}{\mathrm{PV}(0)} = a + \frac{bD}{D + c} \tag{6.3.2}$$

where a, b, c are fitting parameters and $\mathrm{N}(D) = \mathrm{PV}(D)/\mathrm{PV}(0)$ is the normalized film response calculated as the ratio of the read PV for each irradiated strip to the unexposed strip readout. To obtain a specific calibration curve for a given measurement film, two reference strips ideally cut from that film are considered. One of the reference strips remains unexposed while the other is irradiated to a dose 10-30% over the maximum dose in the calculated dose map. Both strips are scanned together with the measurement film and the specific calibration curve is determined from the generic calibration function by linear scaling: $\mathrm{X}(D) = A + BN(D)$, where the parameters A and B can be calculated from the two dose-response pairs $(X_i, D_i)$ obtained from the reference strips. When using more than two recalibration strips, the parameters A and B can be determined by minimizing the quadratic sum of the differences between the predicted and measured values for each reference strip:

$$\sum\nolimits_i (X(D_i) - X_i)^2 \tag{6.4}$$

More recently, two novel recalibration methods have been presented by Ruiz-Morales et al. (2017), the so-called "Parameter Escalation" and "Axis Escalation". According to the authors, these methods are independent of the dose-response function, film response and dependent variable choices. Thus, the calibration function for a given channel is generically defined as $y = f(x; a_i)$ where y represents the film response and x the dose values, or vice-versa and $a_i, i = 1, \ldots, N$ are the fitting parameters. In the recalibration process, M ≥ 2 pairs of film-response and dose values are obtained $(X_j, Y_j),\ j = 1, \ldots, M$. In the Parameter Escalation method, the adjusted calibration curve is given by: $y = f(x; \lambda_i \cdot a_i)$ and the multiplicative factors $\lambda_i, i = 1, \ldots, N$ are obtained by minimizing:

$$\sum_{j=1}^{M} (Y_j - f(X_{j;}\lambda_i \cdot a_i))^2 \tag{6.5}$$

In the Axis Escalation method, three factors, $\lambda_1, \lambda_2, \lambda_3$ are considered to adjust the calibration curves, such that: $y = \lambda_1 \cdot f(\lambda_2 . x; a_i) + \lambda_3$. The $\lambda_1, \lambda_2, \lambda_3$ are again determined by minimizing the quadratic sum of the differences between the recalibration data and the predicted values:

$$\sum_{j=1}^{M} (Y_j - (\lambda_1 \cdot f(\lambda_2 \cdot X_{j;} a_i) + \lambda_3))^2 \tag{6.6}$$

Ruiz-Morales et al. (2017) have reported, in a recent comparison study, an improvement in dose accuracy when using either a control film strip or the aforementioned recalibration methods (Efficient Protocol, Parameter Escalation and Axis Escalation) in comparison with the uncorrected dose estimates. Regarding the differences between the use of recalibration methods and a lot control film piece, the former ones contributed to a consistently better accuracy and reproducibility, with a reported mean improvement of about 1%, depending on the dose range and color channel under study. Indeed, the authors demonstrated that the first approach mitigates inter-scan variability and film response changes due to environmental factors such as temperature, humidity and UV light, although post-exposure darkening and differences in composition between films from the same

lot/box are not entirely compensated for. The recalibration methods, on the contrary, contribute to the mitigation of all these variability effects (Ruiz-Morales et al., 2017).

Ruiz-Morales et al. (2020) went a step further in the study of the recalibration methods by analyzing the impact of the number of recalibration strips and their exposure doses on the level of the achieved accuracy. They concluded that the use of three calibration strips may be advantageous, depending on the employed recalibration method (either Parameter Escalation or Efficient Protocol). The Parameter Escalation method in combination with at least three recalibration strips provided the most accurate dose estimates (Ruiz-Morales et al., 2020).

### 6.2.4 | Dose uncertainty analysis

When implementing a radiochromic film dosimetry system, a method to estimate the uncertainty in measurement of an unknown dose should be established. Some studies (Aldelaijan et al., 2016; Bouchard et al., 2009; Chiu-Tsao and Chan, 2009; Devic, 2011; Devic et al., 2016, 2009, 2005; Ferreira et al., 2009; Garcia and Azorin, 2013; Girard et al., 2012; Lárraga-Gutiérrez et al., 2018; León-Marroquin et al., 2016; Martisíková et al., 2008; Papaconstadopoulos et al., 2014; Saur and Frengen, 2008; Sorriaux et al., 2013; Tamponi et al., 2016, 2015) can be found in the literature on dose uncertainty estimates for single channel dosimetry, using the error propagation analysis as proposed by Devic et al. (2004). According to the authors, the total dose uncertainty $\sigma_{tot}$ for each color channel, can be estimated as the contribution of two main sources, the experimental uncertainty $\sigma_{exp}$ and the fitting uncertainty $\sigma_{fit}$:

$$\sigma_{tot} = \sqrt{\sigma^2{}_{exp} + \sigma^2{}_{fit}} \tag{6.7}$$

The experimental uncertainty results from the uncertainties associated with film irradiation and scanning procedures, including: linac output variations and beam dose calibration, film uniformity, intra-lot variabilities arising from differences in composition/thickness of the sensitive layer of the films, changes in film response due to temperature, UV-light and temporal mismatch, response measurement reproducibility, scanner non-uniform response, and noise. The fitting uncertainty is that associated with the calibration curves fit procedure (Devic et al., 2004). Some comprehensive summaries of the sources of uncertainty that may affect film dosimetry, possible actions to minimize them and expected magnitudes can be found in the literature (Bouchard et al., 2009; Chiu-Tsao and Chan, 2009; Devic et al., 2016; León-Marroquin et al., 2016; Papaconstadopoulos et al., 2014; Sorriaux et al., 2013). The reported total uncertainty values depend on the protocol, the considered sources of uncertainty, on the dose range of interest, color channel and even on the scanner and film model. An overall uncertainty of about 2% (one sigma), for single channel dosimetry (R, G) can be expected.

When using multichannel algorithms for dose determination, the law of propagation of uncertainties may not be easy to implement, as the film responses of the three color channels and their correlations (González-López et al., 2016), as well as the multichannel algorithms particularities should be considered (Vera-Sánchez et al., 2018). Indeed, studies presenting an uncertainty budget for dose determination using multichannel dosimetry are lacking in the literature. To address this apparently unresolved issue, Vera-Sánchez et al. (2018), have proposed a multi-stage model and the application of a Monte Carlo approach that demonstrated to be valid for both single channel and multichannel algorithms. The authors have considered a set of EBT3 films irradiated with doses up to 10 Gy and two scanner models representing different light sources technologies (Epson V800 and Epson 10000 XL) to estimate the uncertainty in dose determination

using single channel (R, G, B) and two triple-channel methods (Mayer et al., 2012; Micke et al., 2011). The reported total uncertainty for the red, green and triple-channel methods was around 2-3% in the entire dose range, except for very low doses (< 1 Gy), which were associated with a higher uncertainty. The blue channel presented the highest uncertainty, as expected. Regarding the scanner model, for Epson V800 scanner, multichannel algorithms were found to provide lower uncertainties than single channel algorithms (R,G), whereas with the Epson 10000 XL scanner, lower uncertainties were verified using single channel dosimetry (R, G), for doses higher than 4 Gy (Vera-Sánchez et al., 2018). Besides its successful implementation, this method does not seem of straightforward use in clinical routine and therefore, users should try to keep all uncertainties as low as possible when performing multichannel dosimetry, as recommended by Devic et al. (2016).

When using commercially available tools, the establishment of a comprehensive uncertainty budget may even be more complicated, as there are many software particularities that may not be controlled and/or known by the users. Almady et al. (2016) for instance, in a comparative study of three software tools (FilmQA Pro, Radiochromic.com and Verisoft) have reported differences up to 2.4% in global and 3.1% in local gamma analysis, for a set of head and neck IMRT/VMAT plans (Almady et al., 2016), but with no information on the achieved precision.

## 6.3 | Summary

Radiochromic film based dosimetry systems are a versatile 2D dosimeter, which can provide accurate absolute dose measurements, based on a comprehensive characterization of the film and scanner responses, and the implementation of a strict dosimetry protocol.

When implementing a film dosimetry system in the clinical practice, there are a few essential questions that the medical physicist team should answer and that are related with both the components and the procedures:

- Which film model, EBT3 or EBT-XD?
- Which scanner model?
- Calibration phase: how many strips and to what irradiation doses to establish a calibration curve? What quantity to use to express the film response to radiation? Which calibration function form?
- Scanning: How to define a scanning protocol?
- Dose determination: single channel or multichannel dosimetry?
- Measurement phase: use a reference unexposed strip or a recalibration method? Which recalibration method? How many recalibration strips?
- Uncertainty analysis: how to establish a comprehensive uncertainty budget? Which uncertainty sources to consider?

The answer to each of these questions was discussed in the present review in the light of the current state of the art. The choice of the film model, either EBT3 or EBT-XD will depend on the dose range of interest. Studies show that EBT3 films perform optimally in the dose range up to 10 Gy, making it more appropriate for the verification of conventional fractionated treatments, while EBT-XD films are more suitable for hypofractionated treatments verification, such as stereotactic radiosurgery and stereotactic body radiotherapy (Palmer et al., 2015b). There are however some groups (Cusumano et al., 2017; Palmer et al., 2015b; Ruggieri et al., 2019; San Miguel et al., 2018; Santos et al., 2019b, Wen et al., 2016) that have used EBT3 films for high dose verifications, using either single channel or triple channel dosimetry methods. Wen et al. (2016) have established a film dosimetry protocol for stereotactic radiosurgery and stereotactic body radiotherapy QA using EBT3 films for doses up to 15 Gy. They proposed the red channel for doses below 10 Gy and the green

channel for higher doses in order to reach an overall accuracy of about 1.5%. Palmer et al. (2015b) in a comparison between EBT3 and EBT-XD films, have reported that both film types may be suitable for high dose radiotherapy verification, although EBT-XD films contributed to improved results, for both single and triple channel dosimetry. Regarding EBT3 films, single channel dosimetry lead to a better agreement between planned and measured doses, in comparison with triple channel dosimetry (98.0% vs 71.0%, global gamma passing rate for 3%/2 mm criteria). These differences can be explained by the shallow gradient and lack of separation of the calibration curves for high doses, which turns into an increased uncertainty in the use of triple channel methods.

Regarding the scanner model, flatbed color document scanners with a large bed are advised for film digitization, as the impact of the lateral scanner response uncertainty can be minimized when films are at a central position, forming a single row in the scanning direction, and at a consistent film orientation.

The dosimetry protocol established by the user will determine the achieved precision and accuracy of the dose measurements for the selected film and the scanner models. The calibration process should be repeated for each new lot of films and the accuracy of the established calibration curves evaluated. For that, additional film strips may be irradiated to known doses in the calibration range, and the differences between the measured and expected values assessed. Eventually, any combination of number of film strips, film response quantity and dose-response function is possible, as long as the choice is based on an uncertainty vs dose error analysis. A formalism for uncertainty estimation can be established based on the work published by Devic et al. (2004). A comprehensive scanning protocol needs to be implemented and followed in both the calibration and the measurement phases, such that optimal scanning parameters are defined including the number of scans necessary to achieve a stabilized scanner response after the warm-up time has elapsed, the number of scans to perform per film, the scanning resolution, and the image pixel depth (Ferreira et al., 2009). The need for a lateral non-uniform response correction has to be carefully assessed, based on the dimensions of the irradiated area, dose of interest, film type (EBT3 or EBT-XD) and applied dosimetry method (either single channel or multichannel). The application of a de-noise technique is highly recommended, namely a 2D adaptive Wiener filter, especially for high resolutions (Vera-Sánchez et al., 2016). Regarding the use of single channel or multichannel dosimetry methods for dose determination, multichannel methods may have improved accuracy over single channel dosimetry, compensating for some sources of uncertainty such as film spatial heterogeneities and scanner non-uniform response. However, a comprehensive assessment of the achieved accuracy and precision when using single channel and multichannel methods should be performed as it may vary depending on the film and scanner models, as well as on the dose range of interest (Dufek et al., 2019). As for the multichannel method to consider, various closed-form solutions have been proposed (Mayer et al., 2012; Méndez et al., 2014; Pérez-Azorín et al., 2014), which can in principle be easily programmed. Other approaches are perhaps conceptually or mathematically more complicated, being available in commercial software (Méndez et al., 2018; Micke et al., 2011). The use of a recalibration method is advised as it allows to meet the immediate conditions and force the calibration into agreement at reference levels, contributing to an improved accuracy and to speed up the processing phase.

To establish a comprehensive uncertainty budget for absolute dose determination, the works published by Devic et al. (2004) and Bouchard et al. (2009), may be used as a guide, even though no considerations on multichannel dosimetry are provided, as discussed.

A summary of the aspects to consider when implementing a film dosimetry system in clinical practice is given in Table 6.3.

**Table 6.3** – Summary of the components to consider and few remarks on how to establish a radiochromic film dosimetry system for dose verification in external beam radiotherapy.

| | |
|---|---|
| **Scanner model** | Flatbed document scanner with preferably a large scanning bed |
| **Film model** | EBT3 or EBT-XD, depending on the dose range of interest |
| **Dosimetry protocol** | All related procedures should be optimized in order to obtain the most accurate results with the lowest associated uncertainty |
| **Calibration phase** | 1. Cut and label the calibration strips<br>2. Irradiate calibration strips to known doses, under reference conditions<br>3. Scan the film strips as per the established protocol<br>4. Image processing – use a de-noising technique, apply a lateral artifact correction method<br>5. Create a calibration curve per channel<br>6. Assess the goodness of the calibration function by performing an uncertainty vs error analysis |
| **Measurement phase** | 1. Cut and label the measurement film to ensure a consistent orientation with the calibration films<br>2. Irradiate the measurement film<br>3. Scan the film together with at least two reference strips to apply a recalibration method (Efficient Protocol, Axis Escalation or Parameter Escalation) or with a lot control film piece<br>4. Use the same scanning protocol as in the calibration phase<br>5. Image processing in the very same way as in the calibration phase<br>6. Determination of the unknown dose using either single channel (red, green) or multichannel methods<br>7. Estimate the associated uncertainty in dose determination |

## 6.4 | Conclusion

Radiochromic film based dosimetry systems composed by Gafchromic™ EBT films and a flatbed document scanner are an accurate, reliable and flexible tool for absolute dose verification in external beam radiotherapy. However, as the film-scanner response can be affected by many factors revisited in this review, the achieved precision and accuracy of the dose measurements, strongly depend on the established procedures, i.e. the dosimetry protocol. The dosimetry protocol should be comprehensive and strictly followed during both the calibration and measurement phases. For that, it is important that the users have a full insight and critical thoughts on every single procedure, and base their choices on an error vs uncertainty analysis. This review intended to serve as a guide in the practical implementation of a film based dosimetry system, highlighting the different issues to be considered and the different available options, based on the actual state of the art.

## 6.5 | Funding

This work was supported by FCT, the Portuguese Foundation for Science and Technology, through a PhD scholarship, reference SFRH/BD/118929/2016.

## 6.6 | Declaration of competing interest

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

## 6.7 | References


Akdeniz, Y., Yegingil, Z., 2019. Evaluation of an efficient calibration protocol for use in radiochromic film dosimetry for photon beams. Radiat. Phys. Chem. 156, 31-37. https://doi.org/10.1016/j.radphyschem.2018.10.002

Aldelaijan, S., Alzorkany, F., Moftah, B., Buzurovic, I., Seuntjens, J., Tomic, N., Devic, S., 2016. Use of a control film piece in radiochromic film dosimetry. Phys. Med. 32, 202–207. https://doi.org/10.1016/j.ejmp.2015.12.004

Aldelaijan, S., Devic, S., 2018. Comparison of dose response functions for EBT3 model GafChromic film dosimetry system. Phys. Med. 49, 112–118. https://doi.org/10.1016/j.ejmp.2018.05.014

Aldelaijan, S., Devic, S., Papaconstadopoulos, P., Bekerat, H., Cormack, R.A., Seuntjens, J., Buzurovic, I.M., 2019. Dose–response linearization in radiochromic film dosimetry based on multichannel normalized pixel value with an integrated spectral correction for scanner response variations. Med. Phys. 46, 5336–5349. https://doi.org/10.1002/mp.13818

Almady, B., Wesolowska, P., Santos, T., Izewska, J., 2016. EP-1511: Gamma analysis: testing scanners and software tools. Radiother. Oncol. 119, S698–S699. https://doi.org/10.1016/S0167-8140(16)32761-X

Almond, P.R., Biggs, P.J., Coursey, B.M., Hanson, W.F., Huq, M.S., Nath, R., Rogers, D.W.O., 1999. AAPM's TG-51 protocol for clinical reference dosimetry of high-energy photon and electron beams. Med. Phys. 26, 1847–1870. https://doi.org/10.1118/1.598691

Anderson, S.E., Grams, M.P., Wan Chan Tseung, H., Furutani, K.M., Beltran, C.J., 2019. A linear relationship for the LET-dependence of Gafchromic EBT3 film in spot-scanning proton therapy. Phys. Med. Biol. 64, 55015. https://doi.org/10.1088/1361-6560/ab0114

Ashland, 2020a. Gafchromic EBT3 specifications [WWW Document].

URL http://www.gafchromic.com/documents/EBT3_Specifications.pdf (accessed 4.9.20).

Ashland, 2020b. Gafchromic EBT-XD specifications [WWW Document].

URL http://www.gafchromic.com/documents/EBTXD_Specifications_Final.pdf (accessed 6.15.20).

Ashland, 2020c. GafchromicTM Dosimetry Tools [WWW Document].

URL http://www.gafchromic.com/documents/PC-11802_Gafchromic_EBT3.pdf (accessed 8.31.20).

Ashland, 2020d. FilmQA Pro User's guide [WWW Document].

URL http://www.gafchromic.com/filmqa-software/filmqapro/index.asp (accessed 6.9.20).

Avanzo, M., Dassie, A., Chandra Acharya, P., Chiovati, P., Pirrone, G., Avigo, C., Barresi, L., Dang Quoc, S., Fiagbedzi, E., Navarria, F., Palazzari, E., Bertola, G., De Paoli, A., Stancanello, J., Sartor, G., 2020. Electron radiotherapy (IOERT) for applications outside of the breast: Dosimetry and influence of tissue inhomogeneities. Phys. Med. 69, 82–89. https://doi.org/10.1016/j.ejmp.2019.12.003

Ayoobian, N., Asl, A.S., Poorbaygi, H., Javanshir, M.R., 2016. Gafchromic film dosimetry of a new HDR 192Ir brachytherapy source. J. Appl. Clin. Med. Phys. 17, 194–205. https://doi.org/10.1120/jacmp.v17i2.6005

Bennie, N., Metcalfe, P., 2016. Practical IMRT QA dosimetry using Gafchromic film: a quick start guide. Australas. Phys. Eng. Sci. Med. 39, 533–545. https://doi.org/10.1007/s13246-016-0443-0

Bouchard, H., Lacroix, F., Beaudoin, G., Carrier, J.-F., Kawrakow, I., 2009. On the characterization and uncertainty analysis of radiochromic film dosimetry. Med. Phys. 36, 1931–1946. https://doi.org/10.1118/1.3121488

Calvo-Ortega, J.-F., Pozo, M., Moragues, S., Casals, J., 2017. Fast protocol for radiochromic film dosimetry using a cloud computing web application. Phys. Med. 39, 1–8. https://doi.org/10.1016/j.ejmp.2017.05.072

Campajola, L., Casolaro, P., Capua, F. Di, 2017. Absolute dose calibration of EBT3 Gafchromic films. J. Instrum. 12, 08015–08015. https://doi.org/10.1088/1748-0221/12/08/p08015

Carrasco, M.A., Perucha, M., Luis, F.J., Baeza, M., Herrador, M., 2013. A comparison between radiochromic EBT2 film model and its predecessor EBT film model. Phys. Med. 29, 412–422. https://doi.org/10.1016/j.ejmp.2012.05.008

Casanova Borca, V., Pasquino, M., Russo, G., Grosso, P., Cante, D., Sciacero, P., Girelli, G., La Porta, M.R., Tofani, S., 2013. Dosimetric characterization and use of GAFCHROMIC EBT3 film for IMRT dose verification. J. Appl. Clin. Med. Phys. 14, 4111. https://doi.org/10.1120/jacmp.v14i2.4111

Casolaro, P., Campajola, L., Capua, F. Di, 2019. The physics of radiochromic process: one calibration equation for all film types. J. Instrum. 14, 08006–08006. https://doi.org/10.1088/1748-0221/14/08/p08006

Chiu-Tsao, S.-T., Chan, M.F., 2009. Photon beam dosimetry in the superficial buildup region using radiochromic EBT film stack. Med. Phys. 36, 2074–2083. https://doi.org/10.1118/1.3125134

Chung, J.P., Oh, S.W., Seong, Y.M., Chun, K.J., Chung, H.-T., 2016. An effective calibration technique for radiochromic films using a single-shot dose distribution in Gamma Knife((R)). Phys. Med. 32, 368–378. https://doi.org/10.1016/j.ejmp.2016.02.001

Colodro, J.F.M., Berná, A.S., Puchades, V.P., Amores, D.R., Baños, M.A., 2017. Volumetric-modulated Arc Therapy Lung Stereotactic Body Radiation Therapy Dosimetric Quality Assurance: A Comparison between Radiochromic Film and Chamber Array. J. Med. Phys. 42, 133–139. https://doi.org/10.4103/jmp.JMP_130_16

Costa, K.C., Gomez, A.M.L., Alonso, T.C., Mourao, A.P., 2017. Radiochromic film calibration for the RQT9 quality beam. Radiat. Phys. Chem. 140, 370–372. https://doi.org/10.1016/j.radphyschem.2017.02.032

Cusumano, D., Fumagalli, M.L., Ghielmetti, F., Rossi, L., Grossi, G., Lanzarotti, R., Fariselli, L., De Martin, E., 2017. Sum signal dosimetry: A new approach for high dose quality assurance with Gafchromic EBT3. J. Appl. Clin. Med. Phys. 18, 181–190. https://doi.org/10.1002/acm2.12045

Cusumano, D., Fumagalli, M.L., Marchetti, M., Fariselli, L., De Martin, E., 2015. Dosimetric verification of stereotactic radiosurgery/stereotactic radiotherapy dose distributions using Gafchromic EBT3. Med. Dosim. 40, 226–231. https://doi.org/10.1016/j.meddos.2015.01.001

Devic, S., 2011. Radiochromic film dosimetry: past, present, and future. Phys. Med. 27, 122–134. https://doi.org/10.1016/j.ejmp.2010.10.001

Devic, S., Aldelaijan, S., Bekerat, H., 2019. Impact of inertia on possible fundamental drawbacks in radiochromic film dosimetry. Phys. Med. 66, 133-134. https://doi.org/10.1016/j.ejmp.2019.08.019

Devic, S., Seuntjens, J., Hegyi, G., Podgorsak, E.B., Soares, C.G., Kirov, A.S., Ali, I., Williamson, J.F., Elizondo, A., 2004. Dosimetric properties of improved GafChromic films for seven different digitizers. Med. Phys. 31, 2392–2401. https://doi.org/10.1118/1.1776691

Devic, S., Seuntjens, J., Sham, E., Podgorsak, E.B., Schmidtlein, C.R., Kirov, A.S., Soares, C.G., 2005. Precise radiochromic film dosimetry using a flat-bed document scanner. Med. Phys. 32, 2245–2253. https://doi.org/10.1118/1.1929253

Devic, S., Tomic, N., Aldelaijan, S., Deblois, F., Seuntjens, J., Chan, M.F., Lewis, D., 2012. Linearization of dose-response curve of the radiochromic film dosimetry system. Med. Phys. 39, 4850–4857. https://doi.org/10.1118/1.4736800

Devic, S., Tomic, N., Lewis, D., 2016. Reference radiochromic film dosimetry: Review of technical aspects. Phys. Med. 32, 541–556. https://doi.org/10.1016/j.ejmp.2016.02.008

Devic, S., Tomic, N., Soares, C.G., Podgorsak, E.B., 2009. Optimizing the dynamic range extension of a radiochromic film dosimetry system. Med. Phys. 36, 429–437. https://doi.org/10.1118/1.3049597

Distefano, G., Lee, J., Jafari, S., Gouldstone, C., Baker, C., Mayles, H., Clark, C.H., 2017. A national dosimetry audit for stereotactic ablative radiotherapy in lung. Radiother. Oncol. 122, 406–410. https://doi.org/10.1016/j.radonc.2016.12.016

Dufek, V., Horakova, I., Koniarova, I., 2019. Comparison of different techniques for evaluation of dose distributions in radiotherapy using radiochromic EBT3 films. Radiat. Prot. Dosimetry 186, 357–361. https://doi.org/10.1093/rpd/ncz231

Ferreira, B.C., Lopes, M.C., Capela, M., 2009. Evaluation of an Epson flatbed scanner to read Gafchromic EBT films for radiation dosimetry. Phys. Med. Biol. 54, 1073–1085. https://doi.org/https://doi.org/10.1088/0031-9155/54/4/017

García-Garduño, O.A., Lárraga-Gutiérrez, J.M., Rodríguez-Villafuerte, M., Martínez-Dávalos, A., Rivera-Montalvo, T., 2016. Effect of correction methods of radiochromic EBT2 films on the

accuracy of IMRT QA. Appl. Radiat. Isot. 107, 121–126. https://doi.org/10.1016/j.apradiso.2015.09.016

Garcia, L.I.R., Azorin, J.F.P., 2013. Improving the calibration of radiochromic films by the use of uncertainties in optical density and dose. Med. Phys. 40, 71726. https://doi.org/10.1118/1.4811238

Gholami, S., Mirzaei, H.R., Jabbary Arfaee, A., Jaberi, R., Nedaie, H.A., Rabi Mahdavi, S., Rajab Bolookat, E., Meigooni, A.S., 2016. Dose distribution verification for GYN brachytherapy using EBT Gafchromic film and TG-43 calculation. Reports Pract. Oncol. Radiother. 21, 480–486. https://doi.org/10.1016/j.rpor.2016.06.005

Girard, F., Bouchard, H., Lacroix, F., 2012. Reference dosimetry using radiochromic film. J. Appl. Clin. Med. Phys. 13, 339–353. https://doi.org/10.1120/jacmp.v13i6.3994

González-López, A., Vera-Sanchez, J.A., Ruiz-Morales, C., 2017. The incidence of the different sources of noise on the uncertainty in radiochromic film dosimetry using single channel and multichannel methods. Phys. Med. Biol. 62, N525–N536. https://doi.org/10.1088/1361-6560/aa8f74

González-López, A., Vera-Sánchez, J.A., Ruiz-Morales, C., 2016. Technical Note: Statistical dependences between channels in radiochromic film readings. Implications in multichannel dosimetry. Med. Phys. 43, 2194. https://doi.org/10.1118/1.4945273

Grams, M.P., Gustafson, J.M., Long, K.M., de los Santos, L.E.F., 2015. Technical Note: Initial characterization of the new EBT-XD Gafchromic film. Med. Phys. 42, 5782–5786. https://doi.org/10.1118/1.4930058

Hodapp, N., 2012. The ICRU Report 83: prescribing, recording and reporting photon-beam intensity-modulated radiation therapy (IMRT). Strahlentherapie und Onkol. 188, 97–99. https://doi.org/10.1007/s00066-011-0015-x

Howard, M., Herman, M., Grams, M., 2020. Methodology for radiochromic film analysis using FilmQA Pro and ImageJ. PLoS One 15, e0233562. https://doi.org/10.1371/journal.pone.0233562

Huang, J.Y., Pulliam, K.B., McKenzie, E.M., Followill, D.S., Kry, S.F., 2014. Effects of spatial resolution and noise on gamma analysis for IMRT QA. J. Appl. Clin. Med. Phys. 15, 4690. https://doi.org/10.1120/jacmp.v15i4.4690

IAEA, 2001. TRS-398: Absorbed Dose Determination in External Beam Radiotherapy, Technical Reports Series. International Atomic Energy Agency, Vienna.

Iqbal, K., Mazhar Iqbal, M., Akram, M., Altaf, S., Buzdar, S.A., 2018. Dosimetric verification and quality assurance for intensity-modulated radiation therapy using Gafchromic EBT3 film. J. Radiother. Pract. 17, 85–95. https://doi.org/10.1017/S1460396917000437

Kern, A., Bäumer, C., Kröninger, K., Mertens, L., Timmermann, B., Walbersloh, J., Wulff, J., 2020. Determination of surface dose in pencil beam scanning proton therapy. Med. Phys. 47, 2277–2288. https://doi.org/10.1002/mp.14086

Khachonkham, S., Dreindl, R., Heilemann, G., Lechner, W., Fuchs, H., Palmans, H., Georg, D., Kuess, P., 2018. Characteristic of EBT-XD and EBT3 radiochromic film dosimetry for photon and proton beams. Phys. Med. Biol. 63, 65007. https://doi.org/10.1088/1361-6560/aab1ee

Khan, M.I., Tahir, M.B., Rafique, M., Iqbal, T., Zulfiqar, S., Zahoor, A., Rehman, J.U., Iqbal, K., Chow, J., 2019. Commissioning and evaluation of a radiochromic EBT3 film dosimetry system. J. Radiother. Pract. 18, 10–15. https://doi.org/10.1017/S1460396918000444

Lárraga-Gutiérrez, J.M., García-Garduño, O.A., Treviño-Palacios, C., Herrera-González, J.A., 2018. Evaluation of a LED-based flatbed document scanner for radiochromic film dosimetry in transmission mode. Phys. Medica 47, 86–91. https://doi.org/10.1016/j.ejmp.2018.02.010

León-Marroquin, E.Y., Herrera González, J.A., Camacho López, M.A., Barajas, J.E.V., García-Garduño, O.A., 2016. Evaluation of the uncertainty in an EBT3 film dosimetry system utilizing net optical density. J. Appl. Clin. Med. Phys. 17, 466–481. https://doi.org/10.1120/jacmp.v17i5.6262

León-Marroquín, E.Y., Lárraga-Gutiérrez, J.M., Herrera-González, J.A., Camacho-López, M.A., Villarreal Barajas, J.E., García-Garduño, O.A., 2018. Investigation of EBT3 radiochromic film's response to humidity. J. Appl. Clin. Med. Phys. 19, 283–290. https://doi.org/10.1002/acm2.12337

León-Marroquín, E.Y., Mulrow, D., Darafsheh, A., Khan, R., 2019a. Response characterization of EBT-XD radiochromic films in megavoltage photon and electron beams. Med. Phys. 46, 4246–4256. https://doi.org/10.1002/mp.13708

León-Marroquín, E.Y., Mulrow, D.J., Khan, R., Darafsheh, A., 2019b. Spectral analysis of the EBT3 radiochromic films for clinical photon and electron beams. Med. Phys. 46, 973–982. https://doi.org/10.1002/mp.13330

Lewis, D., Chan, M.F., 2015. Correcting lateral response artifacts from flatbed scanners for radiochromic film dosimetry. Med. Phys. 42, 416–429. https://doi.org/10.1118/1.4903758

Lewis, D., Devic, S., 2015. Correcting scan-to-scan response variability for a radiochromic film-based reference dosimetry system. Med. Phys. 42, 5692–5701. https://doi.org/10.1118/1.4929563

Lewis, D., Micke, A., Yu, X., Chan, M.F., 2012. An efficient protocol for radiochromic film dosimetry combining calibration and measurement in a single scan. Med. Phys. 39, 6339–50. https://doi.org/10.1118/1.4754797

Lewis, D.F., Chan, M.F., 2016. Technical Note: On GAFChromic EBT-XD film and the lateral response artifact. Med. Phys. 43, 643–649. https://doi.org/10.1118/1.4939226

Li, Y., Chen, L., Zhu, J., Liu, X., 2017. The combination of the error correction methods of GAFCHROMIC EBT3 film. PLoS One 12, e0181958. https://doi.org/10.1371/journal.pone.0181958

Liu, H. W., Gräfe, J., Khan, R., Olivotto, I., Barajas, J.E.V., 2015. Role of in vivo dosimetry with radiochromic films for dose verification during cutaneous radiation therapy. Radiat. Oncol. 10, 12. https://doi.org/10.1186/s13014-014-0325-0

Lopes, M.C., Santos, T., Ventura, T., Capela, M., 2019. Application of the TRS 483 code of practice for reference and relative dosimetry in tomotherapy. Med. Phys. 46, 5799-5806. https://doi.org/10.1002/mp.13855

López-Fidalgo, J., Amo-Salas, M., 2020. Optimal dose calibration in radiotherapy. Radiat. Phys. Chem. 174, 108917. https://doi.org/10.1016/j.radphyschem.2020.108917

Marrazzo, L., Zani, M., Pallotta, S., Arilli, C., Casati, M., Compagnucci, A., Talamonti, C., Bucciolini, M., 2015. GafChromic EBT3 films for patient specific IMRT QA using a multichannel approach. Phys. Med. 31, 1035–1042. https://doi.org/10.1016/j.ejmp.2015.08.010

Martin-Viera Cueto, J.A., Parra Osorio, V., Moreno Saiz, C., Navarro Guirado, F., Casado Villalon, F.J., Galan Montenegro, P., 2015. A universal dose-response curve for radiochromic films. Med. Phys. 42, 221–231. https://doi.org/10.1118/1.4903301

Martisíková, M., Ackermann, B., Jäkel, O., 2008. Analysis of uncertainties in Gafchromic EBT film dosimetry of photon beams. Phys. Med. Biol. 53, 7013–7027. https://doi.org/10.1088/0031-9155/53/24/001

Mayer, R.R., Ma, F., Chen, Y., Miller, R.I., Belard, A., McDonough, J., O'Connell, J.J., 2012. Enhanced dosimetry procedures and assessment for EBT2 radiochromic film. Med. Phys. 39, 2147–2155. https://doi.org/doi:10.1118/1.3694100

Mehrens, H., Taylor, P., Followill, D., Kry, S., 2020. Survey results of 3D-CRT and IMRT quality assurance practice. J. Appl. Clin. Med. Phys. 21, 70-76. https://doi.org/10.1002/acm2.12885

Méndez, I., 2015. Model selection for radiochromic film dosimetry. Phys. Med. Biol. 60, 4089–4104. https://doi.org/10.1088/0031-9155/60/10/4089

Méndez, I., Hartman, V., Hudej, R., Strojnik, A., Casar, B., 2013. Gafchromic EBT2 film dosimetry in reflection mode with a novel plan-based calibration method. Med. Phys. 40, 11720. https://doi.org/10.1118/1.4772075

Méndez, I., Peterlin, P., Hudej, R., Strojnik, A., Casar, B., 2014. On multichannel film dosimetry with channel-independent perturbations. Med. Phys. 41, 11705. https://doi.org/10.1118/1.4845095

Méndez, I., Polsak, A., Hudej, R., Casar, B., 2018. The Multigaussian method: a new approach to mitigating spatial heterogeneities with multichannel radiochromic film dosimetry. Phys. Med. Biol. 63, 175013. https://doi.org/10.1088/1361-6560/aad9c1

Méndez, I., Šljivić, Ž., Hudej, R., Jenko, A., Casar, B., 2016. Grid patterns, spatial inter-scan variations and scanning reading repeatability in radiochromic film dosimetry. Phys. Med. 32, 1072–1081. https://doi.org/https://doi.org/10.1016/j.ejmp.2016.08.003

Menegotti, L., Delana, A., Martignano, A., 2008. Radiochromic film dosimetry with flatbed scanners: A fast and accurate method for dose calibration and uniformity correction with single film exposure. Med. Phys. 35, 3078–3085. https://doi.org/10.1118/1.2936334

Micke, A., Lewis, D.F., Yu, X., 2011. Multichannel film dosimetry with nonuniformity correction. Med. Phys. 38, 2523–2534. https://doi.org/doi:10.1118/1.3576105

Miura, H., Ozawa, S., Hosono, F., Sumida, N., Okazue, T., Yamada, K., Nagata, Y., 2016. Gafchromic EBT-XD film: Dosimetry characterization in high-dose, volumetric-modulated arc therapy. J. Appl. Clin. Med. Phys. 17, 312–322. https://doi.org/10.1120/jacmp.v17i6.6281

Moylan, R., Aland, T., Kairn, T., 2013. Dosimetric accuracy of Gafchromic EBT2 and EBT3 film for in vivo dosimetry. Australas. Phys. Eng. Sci. Med. 36, 331–337. https://doi.org/10.1007/s13246-013-0206-0

Niroomand-Rad, A., Blackwell, C., Coursey, B.M., Gall, K., Galvin, J., McLaughlin, W., Meigooni, A., Nath, R., Rodgers, J., Soares, C., 1998. Radiochromic Film Dosimetry: Recommendations of AAPM Radiation Therapy Committee Task Group 55. Med. Phys. 25, 2093–2115. https://doi.org/10.1118/1.598407

Paelinck, L., Neve, W. De, Wagter, C. De, 2006. Precautions and strategies in using a commercial flatbed scanner for radiochromic film dosimetry. Phys. Med. Biol. 52, 231–242. https://doi.org/10.1088/0031-9155/52/1/015

Palmer, A., Bradley, D., Nisbet, A., 2015a. Evaluation and mitigation of potential errors in radiochromic film dosimetry due to film curvature at scanning. J. Appl. Clin. Med. Phys. 16, 425. https://doi.org/10.1120/jacmp.v16i2.5141

Palmer, A., Dimitriadis, A., Nisbet, A., Clark, C.H., 2015b. Evaluation of Gafchromic EBT-XD film, with comparison to EBT3 film, and application in high dose radiotherapy verification. Phys. Med. Biol. 60, 8741–8752. https://doi.org/10.1088/0031-9155/60/22/8741

Papaconstadopoulos, P., Hegyi, G., Seuntjens, J., Devic, S., 2014. A protocol for EBT3 radiochromic film dosimetry using reflection scanning. Med. Phys. 41, 122101. https://doi.org/10.1118/1.4901308

Park, S., Kang, S.-K., Cheong, K.-H., Hwang, T., Kim, H., Han, T., Lee, M.-Y., Kim, K., Bae, H., Su Kim, H., Han Kim, J., Jae Oh, S., Suh, J.-S., 2012. Variations in dose distribution and optical properties of Gafchromic$^{TM}$ EBT2 film according to scanning mode. Med. Phys. 39, 2524–2535. https://doi.org/doi:10.1118/1.3700731

Pérez-Azorín, J.F., Garcia, L.I.R., Ozcoidi, D.M., Almansa, J.F., 2016. Polarized dosimetry method for Gafchromic EBT3. Phys. Med. 32, 972–980. https://doi.org/10.1016/j.ejmp.2016.06.013

Pérez-Azorín, J.F., Ramos Garcia, L.I., Marti-Climent, J.M., 2014. A method for multichannel dosimetry with EBT3 radiochromic films. Med. Phys. 41, 62101. https://doi.org/10.1118/1.4871622

Poppinga, D., Schoenfeld, A.A., Doerner, K.J., Blanck, O., Harder, D., Poppe, B., 2014. A new correction method serving to eliminate the parabola effect of flatbed scanners used in radiochromic film dosimetry. Med. Phys. 41, 21707. https://doi.org/10.1118/1.4861098

Radiochromic.com, 2020. Radiocromic.com Step by Step Tutorials [WWW Document]. URL https://radiochromic.com/tutorials/tutorials.php (accessed 6.6.20).

Rink, A., Lewis, D.F., Varma, S., Vitkin, I.A., Jaffray, D.A., 2008. Temperature and hydration effects on absorbance spectra and radiation sensitivity of a radiochromic medium. Med. Phys. 35, 4545–4555. https://doi.org/10.1118/1.2975483

Ruggieri, R., Naccarato, S., Mazzola, R., Ricchetti, F., Corradini, S., Fiorentino, A., Alongi, F., 2019. Linac-based radiosurgery for multiple brain metastases: Comparison between two mono-isocenter techniques with multiple non-coplanar arcs. Radiother. Oncol. 132, 70–78. https://doi.org/10.1016/j.radonc.2018.11.014

Ruiz-Morales, C., Antonio Vera-Sanchez, J., Gonzalez-Lopez, A., 2020. Optimizing the recalibration process in radiochromic film dosimetry. Phys. Med. Biol. 65, 15016. https://doi.org/10.1088/1361-6560/ab58dd

Ruiz-Morales, C., Vera-Sánchez, J.A., González-López, A., 2017. On the re-calibration process in radiochromic film dosimetry. Phys. Med. 42, 67–75. https://doi.org/10.1016/j.ejmp.2017.08.013

Rutherford, A., Stevenson, K., Tulk, A., Chalmers, A.J., 2019. Evaluation of four different small animal radiation plans on tumour and normal tissue dosimetry in a glioblastoma mouse model. Br. J. Radiol. 92, 20180469. https://doi.org/10.1259/bjr.20180469

San Miguel, F.J.S., Clemente-Gutiérrez, F., Pérez-Vara, C., 2018. Analysis of different procedures for absolute dosimetry with EBT3 radiochromic film. Biomed. Phys. Eng. Express 4, 65008. https://doi.org/10.1088/2057-1976/aae1cd

Sankar, A., Ayyangar, K.M., Nehru, R.M., Kurup, P.G.G., Murali, V., Enke, C.A., Velmurugan, J., 2006. Comparison of Kodak EDR2 and Gafchromic EBT film for intensity-modulated radiation therapy

dose distribution verification. Med. Dosim. 31, 273–282. https://doi.org/10.1016/j.meddos.2006.06.001

Santos, T., Lopes, M.C., Gershkevitsh, E., Vinagre, F., Faria, D., Carita, L., Pontes, M., Vieira, S., Poli, E., Faustino, S., Ribeiro, F., Trindade, M., Ponte, F., Marcelino, C., Batista, C., Oliveira, S., Figueira, R., Lencart, J., Diaz, E.G., Jacob, K., Brás, S., Pirraco, R., Izewska, J., 2019a. IMRT national audit in Portugal. Phys. Med. 65, 128–136. https://doi.org/10.1016/j.ejmp.2019.08.013

Santos, T., Ventura, T., Capela, M., Mateus, J., Lopes, M.C., 2019b. Influence of film dosimetry protocols on IMRT audit results, in: International Symposium on Standards, Applications and Quality Assurance in Medical Radiation Dosimetry (IDOS 2019). International Atomic Energy Agency, pp. 325–326.

Saur, S., Frengen, J., 2008. GafChromic EBT film dosimetry with flatbed CCD scanner: a novel background correction method and full dose uncertainty analysis. Med. Phys. 35, 3094–3101. https://doi.org/10.1118/1.2938522

Schoenfeld, A.A., Poppinga, D., Harder, D., Doerner, K.-J., Poppe, B., 2014. The artefacts of radiochromic film dosimetry with flatbed scanners and their causation by light scattering from radiation-induced polymers. Phys. Med. Biol. 59, 3575–3597. https://doi.org/10.1088/0031-9155/59/13/3575

Schoenfeld, A.A., Wieker, S., Harder, D., Poppe, B., 2016a. Changes of the optical characteristics of radiochromic films in the transition from EBT3 to EBT-XD films. Phys. Med. Biol. 61, 5426–5442. https://doi.org/10.1088/0031-9155/61/14/5426

Schoenfeld, A.A., Wieker, S., Harder, D., Poppe, B., 2016b. The origin of the flatbed scanner artifacts in radiochromic film dosimetry-key experiments and theoretical descriptions. Phys. Med. Biol. 61, 7704–7724. https://doi.org/10.1088/0031-9155/61/21/7704

Shameem, T., Bennie, N., Butson, M., Thwaites, D., 2020. A comparison between EPSON V700 and EPSON V800 scanners for film dosimetry. Phys. Eng. Sci. Med. 43, 205–212. https://doi.org/10.1007/s13246-019-00837-3

Shimohigashi, Y., Araki, F., Maruyama, M., Nakaguchi, Y., Kuwahara, S., Nagasue, N., Kai, Y., 2015. Evaluation of a single-scan protocol for radiochromic film dosimetry. J. Appl. Clin. Med. Phys. 16, 5226. https://doi.org/10.1120/jacmp.v16i2.5226

Soliman, K., Altimyat, S., Alrushoud, A., Alenezi, A., Alkhorayef, M., 2018. In-Vivo Dosimetry Method for Measuring Peak Surface Dose Using Radiochromic Films during Computed Tomography Scanning of the Sinus. Int. J. Med. Physics, Clin. Eng. Radiat. Oncol. 07, 151–159. https://doi.org/10.4236/ijmpcero.2018.72013

Sorriaux, J., Kacperek, A., Rossomme, S., Lee, J.A., Bertrand, D., Vynckier, S., Sterpin, E., 2013. Evaluation of Gafchromic® EBT3 films characteristics in therapy photon, electron and proton beams. Phys. Medica 29, 599–606. https://doi.org/10.1016/j.ejmp.2012.10.001

Spelleken, E., Crowe, S.B., Sutherland, B., Challens, C., Kairn, T., 2018. Accuracy and efficiency of published film dosimetry techniques using a flat-bed scanner and EBT3 film. Australas. Phys. Eng. Sci. Med. 41, 117–128. https://doi.org/10.1007/s13246-018-0620-4

Tagiling, N., Ab Rashid, R., Azhan, S.N.A., Dollah, N., Geso, M., Rahman, W.N., 2018. Effect of scanning parameters on dose-response of radiochromic films irradiated with photon and electron beams. Heliyon 4, e00864. https://doi.org/10.1016/j.heliyon.2018.e00864

Tamponi, M., Bona, R., Poggiu, A., Marini, P., 2016. A new form of the calibration curve in radiochromic dosimetry. Properties and results. Med. Phys. 43, 4435. https://doi.org/10.1118/1.4954208

Tamponi, M., Bona, R., Poggiu, A., Marini, P., 2015. A practical tool to evaluate dose distributions using radiochromic film in radiation oncology. Phys. Med. 31, 31–36. https://doi.org/10.1016/j.ejmp.2014.07.009

Torres del Río, J., Forastero, C., Moreno-Torres, M., Molina-Matas, M., Martínez-Felipe, A.M., Moreno, Á., López-Peñalver, J.J., Guirado, D., 2019. Use of radiochromic film dosimetry in radiobiology experiments. Radiat. Phys. Chem. 156, 169–173. https://doi.org/10.1016/j.radphyschem.2018.10.008

van Battum, L.J., Huizenga, H., Verdaasdonk, R.M., Heukelom, S., 2016. How flatbed scanners upset accurate film dosimetry. Phys. Med. Biol. 61, 625–649. https://doi.org/10.1088/0031-9155/61/2/625

Vera-Sánchez, J.A., Morales, C.R., López, A.G., 2016. Characterization of noise and digitizer response variability in radiochromic film dosimetry. Impact on treatment verification. Phys. Med. 32, 1167–1174. https://doi.org/10.1016/j.ejmp.2016.08.019

Vera-Sánchez, J.A., Ruiz-Morales, C., Gonzalez-Lopez, A., 2018. Monte Carlo uncertainty analysis of dose estimates in radiochromic film dosimetry with single-channel and multichannel algorithms. Phys. Med. 47, 23–33. https://doi.org/10.1016/j.ejmp.2018.02.006

Wen, N., Lu, S., Kim, J., Qin, Y., Huang, Y., Zhao, B., Liu, C., Chetty, I., 2016. Precise film dosimetry for stereotactic radiosurgery and stereotactic body radiotherapy quality assurance using GafchromicTM EBT3 films. Radiat. Oncol. 11. https://doi.org/10.1186/s13014-016-0709-4

# CHAPTER 7

# A protocol for absolute dose verification of SBRT/SRS treatment plans using Gafchromic™ EBT-XD films

submitted to Physica Medica, November 2020

T Santos[a,b], T Ventura[b], M Capela[b], J Mateus[b], MC Lopes[b]

[a] Physics Department, University of Coimbra, Coimbra, Portugal
[b] Medical Physics Department, IPOCFG, E.P.E., Coimbra, Portugal

## Abstract

**Purpose:** To provide a practical protocol for absolute dose verification of stereotactic body radiotherapy (SBRT) and stereotactic radiosurgery (SRS) treatment plans, based on our clinical experience.

**Methods:** Procedures for film handling, storage, calibration, scanning, image analysis, and conversion to dose are described in detail for a dosimetry system composed by Gafchromic™ EBT-XD films and a flatbed document scanner. Factors that affect the film-scanner response, such as the lateral response artifact, are also revised and accounted for. The accuracy of the adopted methodology was assessed by taking a set of strips irradiated to known doses over the calibration range, in reference conditions. The film response was converted to dose using red channel and triple channel dosimetry. To illustrate the proposed methodology, pre-treatment verification of ten SBRT and SRS helical tomotherapy plans was performed. The agreement between the planned and measured dose distributions was evaluated using global gamma analysis with a criteria of 3%/2 mm 10% threshold (TH) and 2%/2 mm 20% TH.

**Results:** The relative differences between the determined dose from the strips analysis and the expected doses were 0.9±0.6 % for the red channel and 1.1±0.7% for the triple channel method. Regarding the SBRT/SRS plans pre-treatment verification, the gamma passing rates obtained using the single red channel were comparable with those from triple channel dosimetry. The mean gamma passing rates were 99.5±1.0 % vs 99.1±1.6 % (3%/ 2 mm 10% TH) and 98.6±1.8 % vs 97.8±2.1 % (2%/2 mm 20% TH) for red and triple channel dosimetry, respectively.

**Conclusions:** The proposed film dosimetry protocol allows for accurate absolute dose verification of SBRT/SRS treatment plans, applying both single channel and triple channel methods. It may work as a quick start guide for users that intend to implement a film based dosimetry system in clinical practice.

**Keywords:** Film dosimetry protocol, SBRT/SRS treatment plans, pre-treatment QA

## 7.1 | Introduction

Stereotactic radiotherapy, including stereotactic body radiotherapy (SBRT) and stereotactic radiosurgery (SRS) is becoming standard of care for typically small, well-defined and either intracranial or extracranial lesions [1]. Inherent to this technique are small fields and steep dose gradients, as well as increased accuracy and precision requirements, given the high doses delivered in a single or in a few fractions. Pre-treatment quality assurance (QA) verifications should thus be performed to verify the agreement between planned and measured dose distributions [1]. The measurement of such dose distributions is challenging, as it is the selection of a reliable dosimetry system, being the detector resolution a crucial feature. Radiochromic film based dosimetry systems are a potential option as they present a submillimeter spatial resolution, near-water equivalent composition, weak energy dependence in the radiotherapy MV photon energy range, and angular independence [2].

The use of radiochromic film based dosimetry systems is however not widespread [3]. The need for the implementation of a demanding dosimetry protocol may be the main reason. Indeed, the dosimetry protocol should contemplate several aspects, including film calibration and absorbed dose determination. Multiple factors that affect the film-scanner response like post-irradiation darkening, intra-lot variabilities, inter-scan variations, non-uniform response of the scanner, and dependence of the measured response on the film orientation on the scanner bed, should also be taken into account [4]. An AAPM Task Group Report (TG-235) [2] on radiochromic film dosimetry has just been released, providing useful guidelines on its use for a wide range of clinical applications, including photon, electron and proton beams, as well as brachytherapy.

Concurrently, a review paper on film dosimetry for dose verification in external photon beam radiotherapy has been published [5]. The practical aspects to take into account in the characterization of the system components (film and scanner model) and the design of the dosimetry protocol are addressed.

This work builds on the previously published review paper [5] through the presentation of a straightforward and comprehensive film dosimetry protocol for absolute dose verification of SBRT and SRS treatment plans, based on our clinical experience. The processes for film storage, handling, irradiation, scanning, image analysis, and absorbed dose determination are described in detail. Various aspects that affect the film-scanner response are also revised and accounted for. The accuracy of the proposed methodology is evaluated by taking a set of film strips irradiated to known doses and its applicability is illustrated for ten SBRT/SRS treatment plans.

## 7.2 | Materials and Methods

When implementing a film dosimetry system in clinical practice, there are three essential components to consider: the film model, the scanner model and the dosimetry protocol. The choice of the film model depends on the dose range of interest. Radiochromic films commercially designated by Gafchromic™ EBT-XD (Ashland, Inc., Wayne, NJ, USA) present a declared better performance from 0.4 to 40 Gy, being well-suited for SBRT and SRS treatments verification [6]. Flatbed color document scanners with a large bed should be preferred to minimize the impact of the scanner non-uniform response [4,5]. The dosimetry protocol has to be comprehensive and strictly followed during both calibration and measurement phases.

The film dosimetry system used in our department for pre-treatment verification of SBRT and SRS plans is composed by Gafchromic™ EBT-XD films (Ashland, Inc., Wayne, NJ, USA) and an

Epson 10000 XL flatbed document scanner (Seiko Epson Corporation, Japan). The adopted dosimetry protocol is described in detail in the sub-sections below.

### 7.2.1 | Film storage and handling

Films are handled with gloves, kept away from light in a black envelop when not being used and stored at room ambient temperature (20°-25°C), as per TG-235 recommendations [2].

### 7.2.2 | Film scanning protocol

To use a flatbed document scanner in clinical practice it is necessary to characterize its performance and establish a scanning protocol. Important aspects to consider include the color bit depth, spatial resolution, the scanner lamp warm-up, inter-scan variations, noise, orientation effect and scanner non-uniform response [5].

The scanning protocol procedures adopted at our department were established based on a previous published work [7]. Films are digitized in transmission mode, with a color depth of 48 bit (16 bit per color channel – red, green and blue), a spatial resolution of 72 dpi and with all color corrections disabled. To ensure a stabilized scanner response, sixteen consecutive pre-scans are performed after the in-built warm-up time of at least 30 min has elapsed. As the measured response depends on the films position and orientation on the scanner, films are aligned along the central axis, in portrait orientation (25.4 cm edge parallel to scan) [8]. A compression glass plate is placed on their top for flatness, hence mitigating potentially significant dose uncertainties arising from film-to-light distance variation [9,10]. The resulting images are saved in tagged image file format (TIFF) and processed using an in-house developed MATLAB R2010a (Mathworks, Natick, MA, USA) routine. To reduce the amount of noise, which may significantly impact the pre-treatment QA results [11], a two-step procedure is adopted. First, the average of four consecutive scans is taken to reduce temporal noise and second, a Wiener filter is applied to the mean image [4,7,12].

### 7.2.3 | Evaluation of the lateral response artifact

Flatbed document scanners present a non-uniform response in the direction perpendicular to scanning, which induces the so-called lateral response artifact (LRA). LRA is characterized by a reduction of the measured pixel value (PV) towards the lateral edges of the scanner bed [13,14]. The magnitude of the LRA at a given lateral position increases with decreasing PV, i.e. increasing dose, and depends on the scanner and film type. The need for LRA correction should thus be evaluated taking into account the film type, dimensions of the irradiated area, dose level and applied dosimetry method (either single channel or multichannel) [4,5]. The effect tends to be more pronounced for red channel and it may be alleviated by using triple channel dosimetry [8,15].

The LRA for this scanner-film pair was assessed by considering a set of strips ($2.8\times10.15$ cm$^2$) exposed to doses of 0, 2, 5, 8, 12, 16, 20 and 24 Gy, which were digitized at the centre of the scanner (L=0 cm) and then moved to lateral off-axis positions: L=-9, -7, -5, -3, 0, 3, 5, 7, and 9 cm [8]. The ratio between the PV measured at each lateral position to the centre of the scanner PV was calculated.

Correction factors were determined from the performed measurements following the methodology proposed by Lewis and Chan [8]. The authors presented a correction method that relates the response measured for each color channel X of a film irradiated to a known dose D and at a lateral position L, with the response at the centre of the scanner C, by: $Response(C, D, X) =$

$A_{L,X} + B_{L,X} \cdot Response(L, D, X)$, where $A_{L,X}$and $B_{L,X}$ are correction coefficients. These coefficients were calculated from the response of the film strips exposed to 0 and 24 Gy.

To validate the LRA corrections, two film sheets cut to 20.3×20.3 cm$^2$ were irradiated to 10 Gy and 20 Gy with a 20×20 cm$^2$ 6 MV photon open field in a Siemens Oncor Avant-Garde linear accelerator. The films were then scanned in two orthogonal orientations (portrait and landscape), as suggested by Lewis and Chan [8]. The profile in the lateral direction for portrait orientation was extracted from the scanned images before and after applying LRA corrections. The lateral profile after correction was compared with the profile in the scan direction for landscape orientation, which is negligibly affected by scanner non-uniformity [7]. As the measured PV depends on the film orientation, the profiles were normalized at a position corresponding to the centre of the scanner bed.

### 7.2.4 | Calibration phase

Film calibration should be performed for each new lot of films, typically through the irradiation of film strips to known dose under reference conditions, defined by dosimetry protocols such as TRS-398 [16] and TG-51 [17]. In our institution, calibration is done for each new box of films. Nine film strips (2.8×10.15 cm$^2$) cut from a single sheet are irradiated to 0, 1, 2, 3, 5, 10, 15, 20 and 25 Gy in a Tomotherapy HD unit (Accuray Inc., Sunnyvale, CA, USA) at the machine specific reference setup [18]. Calibration film strips are scanned simultaneously to avoid inter-scan variations. After image processing, the average PV over a central region of interest (ROI) of 1×1 cm$^2$ are calculated for each scanner color channel. A calibration curve per channel – red (R), green (G) and blue (B) – is established by fitting a rational function to the data:

$$N(D) = \frac{PV(D)}{PV(0)} = a + \frac{b}{D + c} \tag{7.1}$$

where PV(D) is the mean PV obtained for the film strips irradiated to a dose D, PV(0) the mean PV of the unexposed strip, and a, b, and c are fitting parameters obtained using the nonlinear least-squares minimization method [19].

After establishing the dose-response curves, an error and uncertainty analysis is performed. The error for each strip is calculated as the difference between the expected dose and the dose determined from the application of the calibration curves to the measured response. The total uncertainty $\sigma_{tot}$ in dose determination using the calibration curves is estimated from the contributions of two main sources, the experimental uncertainty $\sigma_{exp}$ and the fitting uncertainty $\sigma_{fit}$, using the error propagation analysis [20]. The experimental component includes factors arising from the reference beam absolute dose calibration, scanner reading process (repeatability and stability), film intra-lot variabilities and variations over the central ROI defined within each calibration strip. The uncertainty associated with reference beam absolute dose calibration in Tomotherapy was estimated in a previous paper [21]. The reproducibility and stability of the scanner response was investigated by digitizing the calibration strips in intervals of 1 h, 48 h and that corresponding to the time a given box is usually used in clinical practice [22]. To analyze intra-lot variabilities, the standard deviation in the response of three strips cut from different films and scanned together was considered.

### 7.2.5 | Measurement phase

Once having the calibration curves per color channel, the response of any irradiated film from the same lot/box can be converted to dose. The time elapsed between irradiation and scanning should be ideally the same for both calibration and measurement films, to match post-irradiation darkening, unless a recalibration protocol is used [19,23]. Film's response conversion to dose can be performed by using single channel or multichannel dosimetry methods. In single channel methods, the calibration curve corresponding to the selected channel is directly applied. Multichannel methods, namely triple channel algorithms, use the information provided by the three color channels to determine a single dose value. Triple channel methods have demonstrated potentially improved dosimetric accuracy over single channel, compensating for variations in the film-scanner response like differences in the thickness of the film active layer, LRA effect, fingerprints, scratches and dust [24–27].

In our department, to analyze a given measurement film, the recalibration method proposed by Lewis et al. [19] is employed. Two reference strips cut from that film are considered (2.8×20.3 $cm^2$), one remains unexposed and the other is irradiated to a dose 10% over the maximum expected dose value in the measurement film. Both strips are scanned together with the measurement film and used to linearly rescale the established calibration curves [19]. This procedure allows for compensation of eventual film-scanner response changes from the calibration. It also contributes to significantly reduce the waiting time to scan the films from days to a few hours [19]. The conversion of film's response to dose is done using single red channel and the triple channel method proposed by Mayer et al. [28], which can be easily programmed in MATLAB.

**Validation of the established protocol**

Before starting using a given film dosimetry protocol for pre-treatment QA, it is important to evaluate its accuracy in reference conditions, where the dose delivered to the film is known. That was done by considering a set of film pieces (2.8×10.15 $cm^2$) irradiated to doses in the range of interest 5, 8, 12, 14, 16, 18, 20, and 22 Gy, in a different day and in the same setup as the calibration films. Two reference strips exposed to 0 and 24 Gy were considered to adjust the calibration curves [19]. The determined dose using red, green, and blue channels and the triple channel method was compared with the film expected dose.

**Pre-treatment verification of SBRT and SRS treatment plans**

The film dosimetry protocol was applied to ten helical tomotherapy plans from patients that underwent SBRT and SRS treatments. The prescription dose varied between 13 and 21 Gy per fraction. All plans were created in the Tomotherapy dedicated treatment planning system (Accuray Inc., Sunnyvale, CA, USA) to be delivered by a Tomotherapy HD unit. For pre-treatment QA verification, the treatment plans were recalculated in the Tomotherapy phantom (Cheese phantom) and a Gafchromic™ EBT-XD film was used to assess the dose distribution in a coronal plane. Simultaneous point dose measurements were also performed using an Exradin A1SL ionization chamber (IC) placed in the Cheese phantom. The dose to each pixel was calculated using both red channel and the triple channel method referred above. The agreement between planned and measured dose distributions is clinically evaluated by using global gamma analysis with a criterion of 3% dose difference relative to the maximum TPS dose in the film plane and 2 mm distance-to-agreement with 10% threshold (TH) [29] in RIT113 software v5.1 (Radiological Imaging Technology Inc., Colorado Springs, CO, USA). The passing rate acceptance limit is 95% [29]. More stringent criteria was also adopted for the purpose of this work, namely 2%/2 mm 20% TH.

For IC measurements, a difference between the planned and measured dose of less than ±3% is considered acceptable in our clinical routine.

The local film dosimetry protocol is summarized in Table 7.1.

**Table 7.1** – Summary of the local film dosimetry protocol.

| | |
|---|---|
| **Film storage and handling** | 1. Films are handled with gloves and kept way from light in a black envelop<br>2. Films are stored at room ambient conditions (20°-25°C) |
| **Scanning protocol** | 1. Color bit depth of 48 bits is used<br>2. Spatial Resolution of 72 dpi is used<br>3. All scanner color corrections are turned-off<br>4. 16 consecutive scans are performed to ensure a stabilized scanner response, after the in-built warm-up time has elapsed<br>5. Films are scanned in a consistent orientation<br>6. Films are aligned along the scanner central axis to minimize the LRA effect<br>7. A glass compression plate is placed on the films top to mitigate film-to-light distance variations<br>8. 4 scans are performed per film<br>9. Images are saved in TIFF format<br>10. To reduce the amount of noise, an average of four consecutive scans is taken and a Wiener filter is applied to the mean image |
| **Calibration phase** | 1. Calibration is repeated for each new box of films<br>2. Nine film strips are irradiated to 0, 1, 2, 3, 4, 5, 10, 15, 20, 25 Gy in reference conditions<br>3. All calibration strips are scanned simultaneously to avoid inter-scan variations<br>4. The average PV over a central ROI of 1×1 $cm^2$ is calculated for each color channel<br>5. A calibration curve per channel is established by fitting a rational function to the normalized response, calculated as the ratio between PV(D) and PV(0)<br>6. An error and uncertainty analysis is performed to assess the goodness of the established calibration curves |
| **Measurement phase** | 1. The recalibration method proposed by Lewis et al.[19] is used to rescale the established calibration curves. For that:<br>- Two film strips are cut from the measurement film<br>- One strip remains unexposed and the other is irradiated immediately before/after the measurement film to a dose 10% over the maximum expected dose in the film plane<br>- The strips are scanned simultaneously with the measurement film<br>2. If needed, LRA effect is taken into account applying the Lewis and Chan [8] correction method<br>3. Red channel and the Mayer's triple channel method are used for dose determination |

## 7.3 | Results

### 7.3.1 | Evaluation of the lateral response artifact

Figure 7.1 shows the ratio between the red channel PV measured for films exposed to 0, 2, 5, 8, 12, 16, 20 and 24 Gy and scanned at a given lateral position (L=-9, -7, -5, -3, 3, 5, 7, 9 cm), and the PV obtained when the films were placed at the centre of the scanner bed. The PV difference between the scanner centre and a 3 cm lateral off-axis position is within 2%, for all dose levels under study. The differences were up to 9% at a 9 cm lateral off-axis displacement. The LRA effect

is less pronounced for green and blue channels, with a maximum difference of 3.7% and 2.8% at 9 cm displacement, respectively.

The LRA correction factors determined using the Lewis and Chan methodology were applied to two films irradiated with 10 Gy and 20 Gy in a 20×20 $cm^2$ field. The differences in PV between the lateral profile obtained after LRA correction were generally within ±1% in comparison with the longitudinal profile in landscape orientation, taken as reference. The profiles from the red channel images corresponding to the film irradiated to 20 Gy are shown in Figure 7.2 in the range for which the correction factors were determined (-9 cm to +9 cm). The lateral profiles before and after applying the LRA correction are represented in cyan and yellow, respectively and the longitudinal profile is given in black. The gray dashed lines correspond to ±1% of the longitudinal pixel values.

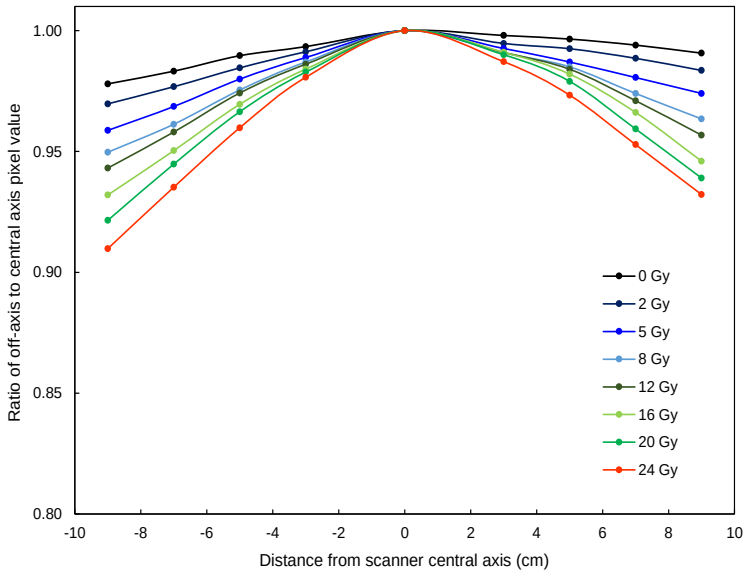


**Figure 7.1** – Lateral response artifact for the dosimetry system composed by EBT-XD films and an Epson 10000 XL scanner in the dose range 0-24 Gy.

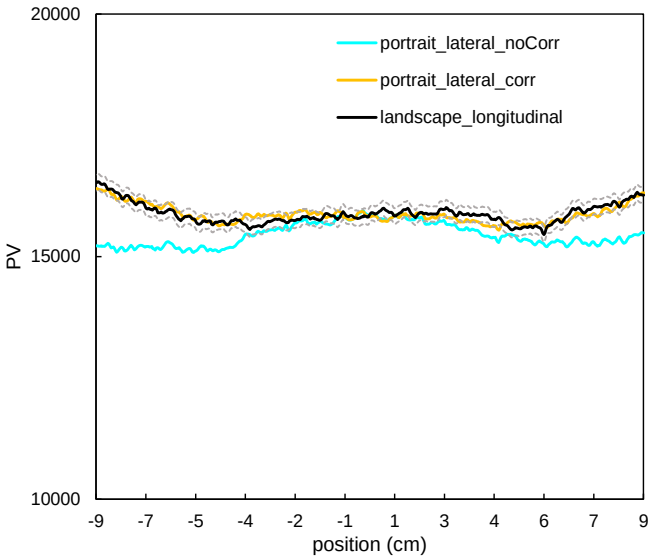


**Figure 7.2** – Profiles obtained for the red channel in the direction perpendicular to scan before LRA correction (cyan), and after LRA correction (yellow) with film in portrait orientation; profile in scan direction with the film digitized in landscape orientation (black). The dashed lines represent ±1% of the longitudinal profile values, taken as reference.

### 7.3.2 | Calibration phase

The calibration curves obtained for the red, green and blue channels over a dose range from 0 to 25 Gy are shown in Figure 7.3. The R-square values, which assess the goodness of the fit, are very close to unity.

The percent difference between the expected dose and the dose obtained from the application of the calibration curves was generally within 2% for all channels, except for low doses (<3 Gy). The mean absolute percent error for the entire dose range was: 1.0±0.5%, 1.6±0.8% and 2.0±1.0% for red, green and blue channels, respectively.

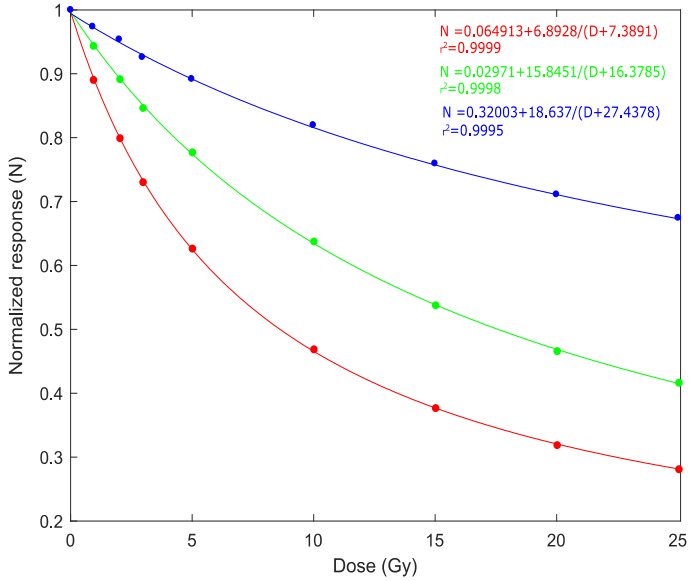


**Figure 7.3** – Dose response curves and corresponding $r^2$ values for the red, green and blue channels.

A summary of dose uncertainties obtained for the red, green and blue channels averaged over the entire dose range is presented in Table 7.2. Scanner non-uniform response and orientation effects were not considered as films were scanned at the centre of the scanner bed, keeping a consistent orientation.

**Table 7.2** – Summary of dose uncertainties in percentage (%) obtained for the red, green and blue channels, over the entire range, with a coverage factor (k) equal to 1.

| **Source of uncertainty** | **Red (%)** | **Green (%)** | **Blue (%)** |
|---|---|---|---|
| Reference absolute dose calibration | 1.1 | 1.1 | 1.1 |
| Reproducibility and stability of the scanner response | 0.2 | 0.2 | 0.3 |
| Film intra-lot variabilities | 0.2 | 0.2 | 0.3 |
| Fluctuations within the 1×1 $cm^2$ ROI | 0.7 | 0.9 | 2.1 |
| Fitting | 0.5 | 0.8 | 1.0 |
| **Total uncertainty (k=1)** | 1.4 | 1.6 | 2.6 |

### 7.3.3 | Measurement phase

**Validation of the established protocol**

The mean absolute percent deviations between the determined doses from the analysis of the validation strips and the expected doses were: 0.9±0.6 % for red channel, 0.9±0.5 % using green channel, 1.4±0.6 % for blue channel and 1.1±0.7% with triple channel dosimetry.

**Pre-treatment verification of SBRT and SRS treatment plans**

The pre-treatment QA results obtained for the ten SBRT/SRS plans, namely the gamma passing rates and ionization chamber percent difference are presented in Table 7.3. The gamma passing rates for red channel were comparable with those from the triple channel dosimetry, when using both 3%/2 mm 10% TH and 2%/2 mm 20% TH criteria. The mean values were 99.5±1.0 % vs 99.1±1.6 % (3%/ 2mm 10% TH) and 98.6±1.8 % vs 97.8±2.1 % (2%/2 mm 20% TH) for the red channel and the triple channel method, respectively. The average difference between measured and calculated doses using IC was 0.7±1.4 %.

**Table 7.3** – Gamma passing rates (%) obtained for the SBRT and SRS treatment plans using red channel and triple channel dosimetry for a criteria of 3%/2 mm 10%TH and 2%/2mm 20%TH, and ionization chamber percent difference (IC %diff).

| | 3%/2 mm 10%TH | | 2%/2 mm 20%TH | | IC %diff |
|---|---|---|---|---|---|
| Case no. | **Red** | **RGB** | **Red** | **RGB** | |
| 1 | 99.5 | 100 | 96.8 | 98.4 | 1.4 |
| 2 | 99.8 | 99.7 | 99.5 | 98.2 | -1.5 |
| 3 | 100 | 95.5 | 100 | 94.2 | 0.4 |
| 4 | 96.6 | 100 | 97.5 | 100 | 0.2 |
| 5 | 99.8 | 96.7 | 94.6 | 95.2 | 1.5 |
| 6 | 99.7 | 100 | 98.3 | 99.6 | 3.0 |
| 7 | 100 | 100 | 99.6 | 98.2 | 0.1 |
| 8 | 100 | 100 | 100 | 100 | 2.5 |
| 9 | 100 | 99.8 | 99.9 | 98.2 | -0.6 |
| 10 | 100 | 98.9 | 99.6 | 95.8 | 0.6 |
| mean±SD | 99.5±1.0 | 99.1±1.6 | 98.6±1.8 | 97.8±2.1 | 0.7±1.4 |

## 7.4 | Discussion

The dosimetry protocol is an essential component of a radiochromic film dosimetry system as it determines the achievable precision and accuracy for the selected film and scanner models. Given the number of factors to be considered, defining reliable and consistent procedures tends to be a cumbersome task. Therefore, the dissemination of comprehensive methodologies clinically used may potentially help new comers to the field. Some authors have already presented more or less detailed dosimetry protocols based on their own clinical experience [30–32]. Howard et al. [31] has just published a methodology for film analysis using a commercially available software, FilmQA Pro, and Image J, a freely available package. It mainly focuses on film calibration procedures and validation of the established dose-response curves using both software solutions. Wen et al. [32] has established a dosimetry protocol for SBRT and SRS treatment verification with Gafchromic™ EBT3 films. In fact, there are some groups that have used EBT3 films for dose verification in hypofractionated regimens, but the low gradient and the lack of separation of the calibration curves lead to an increased uncertainty at doses higher than 10 Gy [33,34]. The film's manufacturer recommends Gafchromic™ EBT-XD films for doses above that (10 Gy) [6].

In this work, a comprehensive film dosimetry protocol for absolute dose verification of SBRT/SRS using Gafchromic™ EBT-XD films is proposed, based on our clinical experience. It describes in detail the adopted procedures for scanning, calibration and absorbed dose determination, which should be strictly followed during both calibration and measurement phases. The accuracy of the proposed methodology was also evaluated by considering a set of film strips irradiated to known doses and its applicability is illustrated for ten SBRT/SRS treatment plans.

Briefly, the scanning protocol should contemplate various factors including the color bit depth, spatial resolution, the scanner lamp warm-up, inter-scan variations, noise, orientation effect and lateral response artifact. The lateral response artifact can be a significant source of error if not assessed and eventually compensated for. Differences in PV for the used film-scanner pair were up to 9% at 9 cm, for 24 Gy. However, given the small size of the lesions treated at our department (maximum dimension < 3 cm) and that the highest dose area is aligned at the centre of the scanner bed, there is generally no need for LRA correction. Still, if larger lesions are to be treated, corrections may be necessary, especially when performing single channel dosimetry [8,15]. Among the LRA correction methods found in the literature [7,9,32,35,36], that proposed by Lewis and Chan [8] has the advantage of being of straightforward implementation, as well as independent of the films production lot and dose.

The calibration should be repeated for each new lot/box of films and the dose error and uncertainty of the established dose-response curves evaluated [2,5]. Using the proposed protocol, the obtained relative percent difference between the measured and expected doses was generally within 2% for red, green and blue channels, over the entire dose range. A dose difference of 2% is deemed acceptable, except for low doses, where higher deviations may be observed. An overall uncertainty of 1.4%, 1.6%, 2.6% (k=1) was estimated for dose determination using the red, green and blue channels corresponding calibration curves, which values are consistent with those reported in the literature [22,37]. Even if absolute uncertainty values should not be directly compared, as different scanners, film models, dosimetry protocols and sources of uncertainty are considered, an overall uncertainty of about 2% for red and green channel dosimetry (k=1) can be expected. The uncertainty associated with triple channel dose determination was not estimated as the law of propagation of uncertainties cannot be easily implemented. For that, it would be necessary to simultaneously account for the film responses of the three color channels, their correlations [38] and the triple channel algorithm particularities [39]. Indeed, studies presenting an uncertainty budget for dose determination using multichannel methods are lacking in the literature. Even the recent comprehensive AAPM report TG-235 [2] does not address this question.

Once having the calibration curves, the measured response for any film from the same lot/box can be converted to dose. To do so, the use of a recalibration method is advised. Recalibration methods allow to compensate for film-scanner response changes from the calibration and forces the calibration into agreement at reference levels, contributing to an improved accuracy and to speed up the processing phase [19,23]. In the proposed protocol, the recalibration method by Lewis et al. [19], which is the most commonly reported in literature, is employed. For dose determination, both single red channel and the triple channel method by Mayer et al. [28] are utilized. The choice of the red channel was based on an analysis of the sensitivity of the dose-response curves given by their first derivate. The first derivate of the red channel calibration curve is higher or almost equal to that of the green channel in the entire dose range of interest, which justifies its selection. Similar conclusions were reported by other authors [34,40,41]. The triple channel method proposed by Mayer et al. [28] can be easily programmed, and it is a closed-form solution of the Micke et al. [24] algorithm implemented in FilmQA Pro (Ashland, Inc., Wayne, NJ, USA). Commercially available packages like FilmQA Pro and Radiochromic.com (web application), may have the advantage of

coming with instructions on how to set-up a film dosimetry system but even so, the users should have a full insight on every single step of the process.

The accuracy of the established film dosimetry system should be evaluated before adopting it for pre-treatment QA verification. That can be done by considering a set of film strips irradiated to known doses in the calibration range. For the film strips irradiated at our institution, differences were within 2% for the majority of the tested doses, using red, green, blue and triple channel dosimetry. It can thus be considered a reliable tool for pre-treatment QA.

Verification results of the ten SBRT/SRS plans confirmed a very good agreement between the calculated and measured dose distributions using both single and triple channel dosimetry. The gamma passing rates obtained for red channel were indeed comparable with those from triple channel dosimetry, using 3%/2 mm 10% TH and 2%/2 mm 20% TH criteria. The selection of the analysis criteria was based on the recent AAPM TG-218 [29] report on methodologies and tolerances limits for pre-treatment QA. This report recommends the use of global gamma analysis with a 3%/2 mm and 10% TH criteria in IMRT plans verification, referring that tighter criteria should be considered for SBRT/SRS techniques. As no specific values are given, 3%/2 mm 10% TH and 2%/2 mm 20% TH were utilized in this study.

While red channel and triple channel dosimetry provide similar results, the triple channel method is advised as it is widely recognized that it allows to compensate for several factors that affect the film-scanner response, including film spatial heterogeneities (variations in the sensitive layer thickness, scratches, fingerprints, dust) [24–27] and scanner lateral response artifact [8,15]. Still, triple channel methods are conceptually and/or mathematically more complicated, preventing their generalized implementation and full understanding. If in-house solutions are adopted, the red channel dosimetry is certainly a reliable alternative as long as a comprehensive approach is followed including the consideration of all the issues, as proposed in the present work.

## 7.5 | Conclusions

A protocol for pre-treatment QA of SBRT and SRS plans using single and triple channel dosimetry was proposed. Procedures for film handling and storage, scanning, calibration, and dose determination were summarized. Factors that may affect the film-scanner response were also revised and considered. Results from verification of the agreement between measured and planned doses for a set of SBRT and SRS treatment plans confirmed the proposed protocol to be an efficient and reliable tool for pre-treatment QA. The described protocol may thus serve as a quick start guide for users that intend to implement film based dosimetry in clinical practice, contributing to its more widespread use. A summary of the local film dosimetry protocol is provided in Table 7.1.

## 7.6 | Funding

This work was supported by FCT, the Portuguese Foundation for Science and Technology, through a PhD scholarship, reference SFRH/BD/118929/2016.

## 7.7 | Declaration of competing interest

None.

## 7.8 | References


[1] Wilke L, Andratschke N, Blanck O, et al. ICRU report 91 on prescribing, recording, and reporting of stereotactic treatments with small photon beams : Statement from the DEGRO/DGMP working group stereotactic radiotherapy and radiosurgery. Strahlenther Onkol. 2019;195(3):193-198. doi:10.1007/s00066-018-1416-x

[2] Niroomand-Rad A, Chiu-Tsao S-T, Grams MP, et al. Radiochromic Film Dosimetry: An Update to TG-55. Med Phys. doi:10.1002/mp.14497

[3] Mehrens H, Taylor P, Followill D, Kry S. Survey results of 3D-CRT and IMRT quality assurance practice. J Appl Clin Med Phys. 21:70-76. 2020. doi:10.1002/acm2.12885

[4] Devic S, Tomic N, Lewis D. Reference radiochromic film dosimetry: Review of technical aspects. Phys Med. 2016;32(4):541-556. doi:10.1016/j.ejmp.2016.02.008

[5] Santos T, Ventura T, Lopes MC. A review on radiochromic film dosimetry for dose verification in high energy photon beams. Radiat Phys Chem. 2021;179:109217. doi:10.1016/j.radphyschem.2020.109217

[6] Ashland. Gafchromic EBT-XD specifications.

http://www.gafchromic.com/documents/EBTXD_Specifications_Final.pdf. Accessed October 21, 2020.

[7] Ferreira BC, Lopes MC, Capela M. Evaluation of an Epson flatbed scanner to read Gafchromic EBT films for radiation dosimetry. Phys Med Biol. 2009;54(4):1073-1085. doi: 10.1088/0031-9155/54/4/017

[8] Lewis D, Chan MF. Correcting lateral response artifacts from flatbed scanners for radiochromic film dosimetry. Med Phys. 2015;42(1):416-429. doi:10.1118/1.4903758

[9] Lewis D, Devic S. Correcting scan-to-scan response variability for a radiochromic film-based reference dosimetry system. Med Phys. 2015;42(10):5692-5701. doi:10.1118/1.4929563

[10] Palmer A, Bradley D, Nisbet A. Evaluation and mitigation of potential errors in radiochromic film dosimetry due to film curvature at scanning. J Appl Clin Med Phys. 2015;16:425. doi:10.1120/jacmp.v16i2.5141

[11] Huang JY, Pulliam KB, McKenzie EM, Followill DS, Kry SF. Effects of spatial resolution and noise on gamma analysis for IMRT QA. J Appl Clin Med Phys. 2014;15(4):4690. doi:10.1120/jacmp.v15i4.4690

[12] Vera-Sánchez JA, Morales CR, López AG. Characterization of noise and digitizer response variability in radiochromic film dosimetry. Impact on treatment verification. Phys Medica. 2016;32(9):1167-1174. doi:10.1016/j.ejmp.2016.08.019

[13] Schoenfeld AA, Wieker S, Harder D, Poppe B. The origin of the flatbed scanner artifacts in radiochromic film dosimetry-key experiments and theoretical descriptions. Phys Med Biol. 2016;61(21):7704-7724. doi:10.1088/0031-9155/61/21/7704

[14] van Battum LJ, Huizenga H, Verdaasdonk RM, Heukelom S. How flatbed scanners upset accurate film dosimetry. Phys Med Biol. 2016;61(2):625-649. doi:10.1088/0031-9155/61/2/625

[15] Lewis DF, Chan MF. Technical Note: On GAFChromic EBT-XD film and the lateral response artifact. Med Phys. 2016;43(2):643-649. doi:10.1118/1.4939226

[16] International Atomic Energy Agency (IAEA). Absorbed Dose Determination in External Beam Radiotherapy. Technical Reports Series No. 398. Vienna: IAEA; 2001

[17] Almond PR, Biggs PJ, Coursey BM, et al. AAPM's TG-51 protocol for clinical reference dosimetry of high-energy photon and electron beams. Med Phys. 1999;26(9):1847-1870. doi:10.1118/1.598691

[18] International Atomic Energy Agency (IAEA). Dosimetry of Small Static Fields Used in External Beam Radiotherapy. Technical Reports Series No. 483. Vienna: IAEA; 2017

[19] Lewis D, Micke A, Yu X, Chan MF. An efficient protocol for radiochromic film dosimetry combining calibration and measurement in a single scan. Med Phys. 2012;39(10):6339-6350. doi:10.1118/1.4754797

[20] Devic S, Seuntjens J, Hegyi G, et al. Dosimetric properties of improved GafChromic films for seven different digitizers. Med Phys. 2004;31(9):2392-2401. doi:10.1118/1.1776691

[21] Lopes M do C, Santos T, Ventura T, Capela M. Application of the TRS 483 code of practice for reference and relative dosimetry in tomotherapy. Med Phys. 2019. doi:10.1002/mp.13855

[22] León-Marroquin EY, Herrera González JA, Camacho López MA, Barajas JEV, García-Garduño OA. Evaluation of the uncertainty in an EBT3 film dosimetry system utilizing net optical density. J Appl Clin Med Phys. 2016;17(5):466-481. doi:10.1120/jacmp.v17i5.6262

[23] Ruiz-Morales C, Vera-Sánchez JA, González-López A. On the re-calibration process in radiochromic film dosimetry. Phys Medica. 2017;42:67-75. doi:10.1016/j.ejmp.2017.08.013

[24] Micke A, Lewis DF, Yu X. Multichannel film dosimetry with nonuniformity correction. Med Phys. 2011;38(5):2523-2534. doi:10.1118/1.3576105

[25] Méndez I, Polsak A, Hudej R, Casar B. The Multigaussian method: a new approach to mitigating spatial heterogeneities with multichannel radiochromic film dosimetry. Phys Med Biol. 2018;63(17):175013. doi:10.1088/1361-6560/aad9c1

[26] Méndez I, Peterlin P, Hudej R, Strojnik A, Casar B. On multichannel film dosimetry with channel-independent perturbations. Med Phys. 2014;41(1):11705. doi:10.1118/1.4845095

[27] Pérez-Azorín JF, Ramos Garcia LI, Marti-Climent JM. A method for multichannel dosimetry with EBT3 radiochromic films. Med Phys. 2014;41(6):62101. doi:10.1118/1.4871622

[28] Mayer RR, Ma F, Chen Y, et al. Enhanced dosimetry procedures and assessment for EBT2 radiochromic film. Med Phys. 2012;39(4):2147-2155. doi:10.1118/1.3694100

[29] Miften M, Olch A, Mihailidis D, et al. Tolerance limits and methodologies for IMRT measurement-based verification QA: Recommendations of AAPM Task Group No. 218. Med Phys. 2018;45(4):e53-e83. doi:10.1002/mp.12810

[30] Bennie N, Metcalfe P. Practical IMRT QA dosimetry using Gafchromic film: a quick start guide. Australas Phys Eng Sci Med. 2016;39(2):533-545. doi:10.1007/s13246-016-0443-0

[31] Howard M, Herman M, Grams M. Methodology for radiochromic film analysis using FilmQA Pro and ImageJ. PLoS One. 2020;15:e0233562. doi:10.1371/journal.pone.0233562

[32] Wen N, Lu S, Kim jg, et al. Precise film dosimetry for stereotactic radiosurgery and stereotactic body radiotherapy quality assurance using GafchromicTM EBT3 films. Radiat Oncol. 2016;11. doi:10.1186/s13014-016-0709-4

[33] Ashland. Gafchromic EBT3 specifications.

http://www.gafchromic.com/documents/EBT3_Specifications.pdf. Accessed November 20, 2020.

[34] Palmer A, Dimitriadis A, Nisbet A, Clark CH. Evaluation of Gafchromic EBT-XD film, with comparison to EBT3 film, and application in high dose radiotherapy verification. Phys Med Biol. 2015;60(22):8741-8752. doi:10.1088/0031-9155/60/22/8741

[35] Poppinga D, Schoenfeld AA, Doerner KJ, Blanck O, Harder D, Poppe B. A new correction method serving to eliminate the parabola effect of flatbed scanners used in radiochromic film dosimetry. Med Phys. 2014;41(2):21707. doi:10.1118/1.4861098

[36] Li Y, Chen L, Zhu J, Liu X. The combination of the error correction methods of GAFCHROMIC EBT3 film. PLoS One. 2017;12(7):e0181958. doi:10.1371/journal.pone.0181958

[37] Tamponi M, Bona R, Poggiu A, Marini P. A new form of the calibration curve in radiochromic dosimetry. Properties and results. Med Phys. 2016;43(7):4435. doi:10.1118/1.4954208

[38] González-López A, Vera-Sánchez JA, Ruiz-Morales C. Technical Note: Statistical dependences between channels in radiochromic film readings. Implications in multichannel dosimetry. Med Phys. 2016;43(5):2194. doi:10.1118/1.4945273

[39] Vera-Sánchez JA, Ruiz-Morales C, Gonzalez-Lopez A. Monte Carlo uncertainty analysis of dose estimates in radiochromic film dosimetry with single-channel and multichannel algorithms. Phys Med. 2018;47:23-33. doi:10.1016/j.ejmp.2018.02.006

[40] Khachonkham S, Dreindl R, Heilemann G, et al. Characteristic of EBT-XD and EBT3 radiochromic film dosimetry for photon and proton beams. Phys Med Biol. 2018;63(6):65007. doi:10.1088/1361-6560/aab1ee

[41] Miura H, Ozawa S, Hosono F, et al. Gafchromic EBT-XD film: Dosimetry characterization in high-dose, volumetric-modulated arc therapy. J Appl Clin Med Phys. 2016;17(6):312-322. doi:10.1120/jacmp.v17i6.6281

# CHAPTER 8

# Conclusions

The present work had as main objectives to conduct the IAEA supported national audit for IMRT dose delivery in Portugal and to contribute to extend the adopted methodology to helical tomotherapy, to include quality and complexity metrics for plan assessment and to establish a practical guide for radiochromic film dosimetry.

The dosimetry audit was carried out in Portugal in 2018 and covered 100% of radiotherapy institutions performing IMRT treatments. Both the Portuguese Physics Society and the Portuguese Heath Authority endorsed the project. Portugal was one of the first countries that have implemented this audit at the national level. Even if some suboptimal parameters were identified, the status of TPS calculations and dose delivery was generally within the specified tolerances, with no major reasons for concern. Adaptations for HT included the definition of procedures to check the machine beam output, the TPS beam model and some small field parameters, in the view of the IAEA/AAPM code of practice TRS 483.

The quality and complexity of the audit plans were evaluated and compared for the different groups of technologies and IMRT/VMAT treatment techniques. Plan quality was assessed using a software tool called SPIDERplan. This tool allows to obtain a global plan score based on the accomplishment of the individual PTVs and organs-at-risk (OARs) dosimetric goals, which is completely independent of the treatment technique and technology. For complexity analysis, a set of metrics that evaluate different plan features were considered. A statistical approach based on principal component analysis was followed with the ultimate aim of computing a global plan complexity score. This global score combines the multiple evaluated aspects into a single numerical indicator, allowing to summarize the differences between and within the groups of audit plans. As no complexity metrics had been proposed in the literature for helical tomotherapy plans, the complexity analysis was extended to HT as a novel contribution through the adaptation and the definition of new indices. The presented methodology for quality and complexity analysis can be applied to any given plans set, which may help to benchmark institutions with similar technology/techniques in the context of IMRT audits.

A review on radiochromic film dosimetry was done with the intent to serve as a guide in the practical implementation of film dosimetry, both in clinical practice or for audit purposes. The different aspects to be considered were discussed in the light of the actual state of the art. Moreover, a straightforward dosimetry protocol for high dose verification of stereotactic treatments was provided, which can eventually be considered in future extensions of the audit methodology to hypofractionated regimens.

Dosimetry audits are a privileged tool to improve treatment quality and patient safety in radiotherapy. However, audit programs can be pricey, time consuming, and the funding for that is scarcely considered. With treatment techniques and technologies becoming more and more complex, a broader access to cost effective and efficient audit programs is needed and completely in line with the European priorities concerning the implementation of the basic safety standards directive 59/2013. The European Commission has recently promoted a new project called QuADRANT that aims to warrant the compliance with this directive and support the implementation of clinical audits in radiology, radiotherapy and nuclear medicine.

The presented doctoral thesis was certainly of great important for the radiotherapy in Portugal, paving the way for future audit projects.

# APPENDIX I

# Application of the TRS 483 code of practice for reference and relative dosimetry in Tomotherapy

Medical Physics, 2019, volume 46, issue 12, pages 5799-5806

MC Lopes[a], T Santos[a,b], T Ventura[a], M Capela[a]

[a] Medical Physics Department, IPOCFG, E.P.E., Coimbra, Portugal
[b] Physics Department, University of Coimbra, Coimbra, Portugal

## Abstract

**Purpose:** To apply the recent code of practice from the IAEA/AAPM, TRS 483, to Helical Tomotherapy (HT) for reference and relative dosimetry obtaining correction factors for the Exradin A1SL ionization chamber.

**Methods:** The beam quality correction factor for the A1SL chamber was obtained through three different approaches following TRS 483 concepts and compared with published values. The determination of the reference absolute dose for the machine specific reference field (msr) was complemented with relative dosimetry through the determination of output factors of small fields using different detectors. The response of A1SL was compared with correction-free film results and corrected output factors of other detectors.

**Results:** A weighted mean beam quality correction factor of 0.9945±0.0073 was obtained for the A1SL chamber which is in agreement with values reported in the literature. Output factors obtained with different detectors were in agreement, given the uncertainty level. Considering the film output factors as free of corrections, the average value for A1SL output factors corrections was 1.000±0.007.

**Conclusions:** The beam quality correction factors for the A1SL chamber obtained through the three different pathways recommended by TRS 483 agreed with each other and also with published values. The measurements from the A1SL chamber normalized to the machine specific reference field in HT can be taken as output factors for small clinical field sizes without further corrections.

**Keywords:** Helical Tomotherapy; TRS 483; reference dosimetry; small field dosimetry.

## A1.1 | Introduction

The joint IAEA and AAPM international code of practice (CoP) for small static fields dosimetry, TRS 483, [1] was issued in December 2017 and the respective summary paper in 2018 [2]. This is, a decade after the publication of a formalism that outlined the extension of the dosimetry based on absorbed dose to water to small and composite fields and to non-reference conditions [3]. Both publications rely on the universally adopted CoP such as the IAEA publication TRS 398 [4] and AAPM TG-51[5]. The established reference dosimetry for conventional 10×10 $cm^2$ fields is extended to small fields and to non-reference fields by introducing the concept of machine specific reference (msr) field. This concept applies to treatment units where the conventional reference field is not permitted and Helical Tomotherapy (HT) is one of such machines. HT is a pre-commissioned equipment as it comes with a gold standard beam model that only needs to be verified by the customer as part of the commissioning process. With the aim of supporting medical physicists in this task, a comprehensive AAPM report (Task Group 148) [6] was published in 2010.

An HT machine was installed in our hospital in February 2016. Two Exradin A1SL ionization chambers, from Standard Imaging (Middleton, WI, USA), are included in the standard dosimetry package for HT. At the time of the commissioning, TRS 483 had not yet been issued and the acceptance tests for HT do not include the concept of reference dosimetry. As the HT unit operates clinically with a rotating beam, the output calibration in rotational mode is verified during the acceptance tests. The corresponding helical IMRT reference plan intends to deliver a uniform 2 Gy dose to a cylindrical structure centrally placed inside the TomoTherapy Phantom (Cheese phantom). This constitutes a plan-class specific reference (pcsr) field according to the Alfonso et al. formalism [3]. TG 148 [6] also recommends such a plan for rotational output verification. The difference between the measurement performed with an A1SL chamber placed at the centre of the defined structure in the Cheese phantom and the HT dedicated treatment planning system (TPS) calculation ignores the quality difference from calibration at the standards laboratory (ADCL) for the chamber response, as per Accuray protocol.

The characteristics of the A1SL chamber in terms of sensitivity, long and short-term stability, ion recombination and polarity, leakage and settling make it suitable for reference dosimetry purposes [7]. Nevertheless, it was not included in the reference dosimeters listed in TRS 483, except for Gamma Knife.

Since the publication of the Alfonso et al. formalism [3], different approaches were followed where the correction factors for the A1SL chamber were obtained with the common aim of performing reference dosimetry in HT with this detector [8–13]. Thomas et al. [8] made the first attempt to obtain from Monte Carlo (MC) simulations the $\%ddx_{[HT\ TG\text{-}51]}$, a dosimetric equivalent in HT of the beam quality index defined in TG-51 [5] as the percentage depth dose at 10 cm depth in a water phantom for a field size of 10×10 $cm^2$ at an SSD of 100 cm, due to photons only. A third order polynomial was fit between $\%ddx_{[HT\ TG\text{-}51]}$ and $\%ddx_{[HT\ Ref]}$, where [HT Ref] refers to the 5×10 $cm^2$ machine specific reference field size in HT, for SSD=85 cm. This polynomial conversion fit is just valid within a limited range of $\%ddx_{[HT\ Ref]}$ between 58.8% and 60.8%. Since the A1SL quality correction factor, $k_Q$ is not included in TG-51 [5], Thomas et al. presented the $k_Q$ graph for A1SL.

In 2011, Zeverino et al. [10] used the virtual water HT slab phantom to measure beam quality specifiers under HT msr conditions with A1SL and two other ionization chambers. They obtained a $\%ddx_{[HT\ Ref]}$ of 60.6%, within the valid range of the Thomas et al. conversion. They obtained a value of 0.998±0.001 for the beam quality correction factor of A1SL out of the Thomas et al. paper [8]. Alternatively, they calibrated the three chambers against Gafchromic EBT2 films, following their local protocol for absolute dose calibration, which only uses the red channel response and a double

exposure of the film. The beam quality correction factor obtained by this alternative calibration was 0.997±0.010.

Using an alternative reference dosimetry based on alanine, two groups [11,12] published correction factors for the A1SL chamber. Ost et al. [11] within the framework of a pilot study involving HT beam output measurements in four centers, derived the beam quality correction factor for A1SL against alanine/EPR dosimetry. Gago Arias et al. [12] also used alanine for independent calibration. Considering a $k_Q$=0.9994 from Ref 8, they obtained a beam quality correction factor for the A1SL chamber in line with Ost et al.

After the issue of TRS 483, the motivation for the present work was to follow the new recommendations exploring the different pathways and concepts embodied in the CoP, in order to obtain the beam quality correction factor for the A1SL chamber in the context of performing reference dosimetry in Tomotherapy. Subsequently, we aimed to validate the Accuray practice of not correcting the chamber reading for quality to verify the rotational output in HT.

In complement, to underpin small field dosimetry in HT, an intercomparison of output factors was also carried. A set of small fields were configured through the pneumatically controlled binary multileaf collimator (MLC), using different detectors like a microDiamond, an unshielded diode, a micro chamber as well as film dosimetry, to compare with the A1SL chamber response. Once again, the information conveyed in TRS 483 has been explored in order to find a sustained way of performing relative dosimetry in HT.

## A1.2 | Materials and Methods

### A1.2.1 | Reference dosimetry – msr

According to both TRS 483 [1] and TG-148 [6] the reference dosimetry setup in the HT unit was defined as corresponding to the absorbed dose to water determined at the reference point $z_{ref}$ located at 10 cm depth in water for a 5×10 cm$^2$ static field size defined at a source surface distance (SSD) of 85 cm. This corresponds to the $f_{msr}$ (machine specific reference field). From the TRS 483 recommendations, three different approaches have been followed to reach a beam quality correction factor for A1SL chamber to work around the lack of a tabulated value for it in the recent CoP: 1) film independent dosimetry; 2) cross-calibration using reference chambers; 3) the hypothetical reference 10×10 cm$^2$ field.

**i) Independent dosimetry approach**

Using an independent dosimetry method like film dosimetry to determine $D_{w,Q_{msr}}^{f_{msr}}(z_{ref})$, the beam quality correction factor for the A1SL ionization chamber, $k_{Q_{msr},Q_0}^{f_{msr},f_{ref}}$ can be obtained through:

$$D_{w,Q_{msr}}^{f_{msr}}(z_{ref}) = M_{Q_{msr}}^{f_{msr}} N_{D,w,Q_0}^{f_{ref}} k_{Q_{msr},Q_0}^{f_{msr},f_{ref}} \quad \text{(A1.1)}$$

where $N_{D,w,Q_0}^{f_{ref}}$ is the chamber calibration coefficient from the standards laboratory certificate, for the conventional 10×10 cm$^2$ reference field at reference beam quality $Q_0$.

For this first approach, a well-established film dosimetry protocol based on an in-house software programmed on Matlab R2010a environment [14,15], using multichannel correction and based on the Mayer et al. methodology [16] was used to determine absorbed dose to water at machine specific reference conditions. The Gafchromic EBT3 film calibration curve was obtained through

the irradiation of 8 film strips of 2.8×10.2 cm$^2$ with known doses, from 0 to 3.5 Gy, using a Farmer PTW-30013 reference chamber, at reference conditions for a 6 MV photon beam in a Siemens Oncor Avant-Garde linear accelerator (linac). Three film strips with 5×6.8 cm$^2$ from the same batch and box were then irradiated in Tomotherapy at the msr setup, assuming film energy independence [17-20] and tracking HT irradiation stability through independent monitoring.

For the film msr setup, the same arrangement as for msr in water was repeated for a virtual water slab phantom, respecting the need for at least 5 cm of backscattering material [1,21] and enabling an estimate for the water/virtual water correction factor.

For film reading, the three film strips were scanned in transmission mode on a flatbed scanner Epson Expression 10000XL (Seiko Epson Corporation) with a colour depth of 48 bits, and a spatial resolution of 72 dpi, and saved in tagged image file format. All colour correction enhancements options were disabled. To ensure a stabilized response, sixteen consecutive empty scans were performed after a warm-up time of at least 30 minutes. The film strips were placed in a central region of the scanner and perpendicularly to the scanning direction. To minimize any film curling a transparent glass was placed on top of the film surface just before the scanning procedure. All film strips were handled with gloves and kept from visible light in black envelopes supplied with the film box until the irradiation or the scanner reading. Each of the three film strips has been scanned four times both individually and simultaneously 24 h after irradiation, in order to account for scanning uncertainty.

For film processing, image conversion from optical density to dose was performed applying the specific calibration function (updated through independent specific film irradiation at the time of film strip irradiation in HT [16]) to a region of interest of 1 cm$^2$ located centrally at each film strip. All dose maps were saved and imported to RIT113 Dosimetry System software v.5.1 (Radiological Imaging Technology, Colorado Springs, CO, USA) where dose profiles were analysed.

**ii) Cross-calibration approach**

According to TRS 483, eq. (A1.1) can be used when the ionization chamber calibrated at the standards laboratory with a $N_{D,w,Q_0}^{f_{ref}}$ calibration coefficient, has a generic beam quality correction factor $k_{Q_{msr},Q_0}^{f_{msr},f_{ref}}$ tabulated in the CoP. Unfortunately, this is not the case for A1SL chamber. However, Table 13 in TRS 483 [1] provides values of this beam quality correction for different reference detectors in flattening filter free (FFF) beams, including Tomotherapy. The Farmer chamber PTW-30013 and the Semiflex 0.3 cc PTW-31013 are two detectors belonging to our dosimetry equipment at the hospital. So, using the cross-calibration expression:

$$\left[N_{Dw,Q_{msr}}^{f_{msr}}\right]_{A1SL} = \frac{\left[D_{w,Q_{msr}}^{f_{msr}}\right]_{REF}}{\left[M_{Q_{msr}}^{f_{msr}}\right]_{A1SL}} \tag{A1.2}$$

where

$$\left[D_{w,Q_{msr}}^{f_{msr}}\right]_{REF} = M_{Q_{msr}}^{f_{msr}} N_{Dw,Q_0}^{f_{ref}} \left(k_{Q_{msr},Q_0}^{f_{msr},f_{ref}}\right)_{REF} \tag{A1.3}$$

With $\left(k_{Q_{msr},Q_0}^{f_{msr},f_{ref}}\right)_{REF}$ taken from Table 13, TRS 483 [1] for the two existing chambers and equating:

$$k_{Q_{msr},Q_0}^{f_{msr},f_{ref}} = \frac{\left[N_{Dw,Q_{msr}}^{f_{msr}}\right]_{A1SL}}{\left[N_{D,w,Q_0}^{f_{ref}}\right]_{A1SL}} \tag{A1.4}$$

the corresponding beam quality correction factors for the A1SL chamber were obtained.

**iii) Hypothetical reference field approach**

If a generic beam quality correction factor does not exist in TRS 483 [1], as it is the case for A1SL chamber, eq. (A1.1) can be further broken down into another term as:

$$D_{w,Q_{msr}}^{f_{msr}}(z_{ref}) = M_{Q_{msr}}^{f_{msr}} N_{D,w,Q_0}^{f_{ref}} k_{Q,Q_0,}^{f_{ref}} k_{Q_{msr},Q}^{f_{msr},f_{ref}} \quad \text{(A1.5)}$$

where $k_{Q,Q_0,}^{f_{ref}}$ is the quality correction factor for an "hypothetical" 10×10 cm$^2$ reference field at a beam quality Q. In TRS 483 the correction from Q to $Q_{msr}$ for the $f_{msr}$, expressed by the factor $k_{Q_{msr},Q}^{f_{msr},f_{ref}}$, is assumed to be unity within the involved uncertainties.

To obtain $k_{Q,Q_0,}^{f_{ref}}$ the equivalent square field size (S) to $f_{msr}$ in HT was taken as 6.6 cm from Table 16 in TRS 483 [1]. The Palmans et al. [22] expression was thus used to derive the quality index for the "hypothetical" 10×10 cm$^2$ reference field in HT as:

$$TPR_{20,10}(10) = \frac{TPR_{20,10}^{HT} + c(10 - S)}{1 + c(10 - S)} \quad \text{(A1.6)}$$

with c=(16.15±0.12) ×10$^{-3}$. $TPR_{20,10}^{HT}$, the tissue phantom ratio for the $f_{msr}$ in HT, was measured taking the ratio of measurements at 20 and 10 cm depths for a constant source detector distance, SDD=85 cm in water using the A1SL chamber. As this is a quality index applicable in the framework of TRS 398 [4], the corresponding beam quality correction factor, $k_Q$ (coincident with $k_{Q,Q_0,}^{f_{ref}}$) could be used. The $k_Q$ for this chamber is tabulated in TRS 398 [4].

For all ionometric measurements, in each of the three approaches, a careful assessment of all charge collection influencing quantities as temperature, air pressure, polarity, and ion recombination has been done. The corresponding associated uncertainties were estimated. Also, detector in-water positioning errors and beam stability have been accounted for in the overall uncertainty associated with the determination of reference absolute dose for *msr* field in HT.

### A1.2.2 | Relative dosimetry – output factors in HT

Output correction factors for Tomotherapy are tabulated in Table 24, TRS 483 [1] for a list of suitable detectors to perform small field dosimetry. Those field size dependent corrections, $k_{Q_{clin}Q_{msr}}^{f_{clin}f_{msr}}$ account for radiation features characterizing small fields and also for intrinsic characteristics of the detectors. So the output factor for a given small square clinical field, $S_{clin}$, is obtained correcting the measurements ratio between $S_{clin}$ and $f_{msr}$ by $k_{Q_{clin}Q_{msr}}^{f_{clin}f_{msr}}$:

$$\Omega_{Q_{clin}Q_{msr}}^{f_{clin}f_{msr}} = \frac{M_{Q_{clin}}^{f_{clin}}}{M_{Q_{msr}}^{f_{msr}}} k_{Q_{clin}Q_{msr}}^{f_{clin}f_{msr}} \quad \text{(A1.7)}$$

Table 24 in TRS 483 [1] reports those correction factors for different detectors and a range $S_{clin}$ from 0.4 to 8 cm for Tomotherapy machines. The PTW-31016 PinPoint 3D, the 60018 unshielded diode and the 60019 microDiamond appear in the referred table. All these detectors have been used in the present study, for the measurement of output factors in HT.

A set of static irradiation procedures in our HT unit for each of the three field sizes as defined by the Y jaws – 1, 2.5 and 5 cm – opening successively 2, 4, 6, 8, 12, and 16 central MLC leaves have been created. These settings produced, for each of the defined Y dimension, field widths at SSD=85cm of 1.25, 2.5, 3.75, 5, 7.5 and 10 cm.

The setup for these measurements was SSD=85 cm and the detectors were placed with their reference point at 10 cm depth in water, according to the alignment recommendations in TRS 483 [1]. Again following the recommendations concerning small fields [1], the field sizes at the measurement depth have been estimated through the measured profiles FWHM, for Y. For X, as the 50% of maximum dose may be located within the in-field region of the FFF profile, the measured width may not correspond to the correct field size and in HT the normalization methods proposed for FFF beams for conventional linear accelerators [23,24] cannot be used. The field size for X was thus obtained multiplying the nominal width by the scale factor of 95/85 accounting for the setup. In HT, the pneumatic binary MLC leaves are either completely closed or open. This way, for the created static beams, the field widths in X are uniquely determined by geometry.

Measuring field size dimensions can be critical in small fields because of the associated effects [3], so care should be taken on the choice of the detector as well as its alignment. For the configured small fields in HT, FWHM of the longitudinal profiles (Y dimension) at 10 cm depth with SSD=85 cm were measured with the PTW-60019 microDiamond detector. With a sensitive volume of 0.004 $mm^3$, this detector was considered small enough. The detector was aligned parallel to the beam axis in the motorized PTW MP3-T water phantom.

To use Table 24 in TRS 483 [1] the equivalent square field size for each of the configured HT small fields was calculated. Given the uneven in-plane (Y) and cross-plane (X) field dimensions, the equivalent small field sizes have been calculated by the geometric mean $S_{clin} = \sqrt{XY}$ even if some of the field sizes are outside the condition 0.7<Y/X<1.4. As referred in TRS 483 [1], for those cases a larger uncertainty on the output correction factor has to be considered.

Output factors have been determined for a PTW-31016 PinPoint 3D, a 60018 unshielded diode and a 60019 microDiamond, interpolating the field size dependent correction factors $k_{Q_{clin}Q_{msr}}^{f_{clin}f_{msr}}$ included in Table 24, TRS 483 [1]. Each detector was aligned parallel (or perpendicular) to the beam axis. Proper central axis correction was applied using the center check module of the beam data acquisition software PTW-MEPHYSTO mcc 3.2. As an independent dosimetry free of correction, output factors for the same field sizes have been measured using Gafchromic EBT3 film dosimetry, following the local protocol referred above. Finally, the A1SL chamber output factors for the same field sizes range were determined through measurements at the same depth with the chamber axis perpendicular to the beam axis. Comparison with the results of film dosimetry was done to infer the possible need for field size dependent output correction factors for A1SL chamber.

## A1.3 | Results

### A1.3.1 | Reference dosimetry – msr

Table A1.1 summarizes the results from present work and from literature for the different approaches used to assess the beam quality correction factor for the A1SL Exradin ionization chamber for reference dosimetry in the HT *msr* field. Values of the beam quality correction factor ranging from 0.9900 to 0.9988 were found in the present work and from 0.9814 to 0.9973 in the literature, with different associated uncertainties.

**Table A1.1** – Beam quality correction factor, $k_{Q_{msr}Q_0}^{f_{msr},f_{ref}}$ for the Exradin A1SL chamber obtained by different approaches in this work and published in the literature. The average values were calculated as the weighted mean of the values obtained in this work and unweighted mean for those in the literature.

| Present work | | From literature | | Method |
|---|---|---|---|---|
| Reference Film dosimetry | 0.9978±0.0162 | Thomas *et al.* [8] | 0.9973±0.002 | Monte Carlo |
| Cross-calibration (Farmer) | 0.9900±0.0129 | Zeverino *et al.* [10] | 0.997±0.010 | Film |
| Cross-calibration (PTW-31013) | 0.9935±0.0149 | Ost *et al.* [11] | 0.996±0.012 | Alanine |
| Through $TPR_{20,10}$ (10) | 0.9988±0.0150 | Sterpin *et al.* [9] | 0.997±0.002 | Monte Carlo |
| | | Gago Arias *et al.*[2] | 0.9814±0.019 | Alanine |
| **Mean value** | **0.9945±0.0073**[a] | **Mean value** | **0.9937±0.0046**[b] | |

[a] weighted mean with corresponding associated uncertainty

[b] unweighted mean with estimated uncertainty considering a rectangular distribution over the publish values

In the first approach followed in this work, the reported uncertainty (k=2) of 1.6% for film dosimetry combines the experimental uncertainty, which includes the contributions of scanner reading process (0.2%), the response of different films from the same lot (0.1%), and absolute dose calibration (0.6%), with the uncertainty due to the fit of the calibration curve which is around 0.5% for the considered dose range [17].

From the cross-calibrations approach, the main uncertainty associated with each of the two derived values is due to the corresponding expanded combined uncertainty of each calibration coefficient from the corresponding certificate, with an additional 0.2% due to the correlations inherent to the cross-calibration procedure [4].

In the third approach, using the virtual water slab phantom, and after having confirmed its equivalence to water within 0.4% (as referred also by Ost *et al.* [11]), a $TPR_{20,10}^{HT}$ of 0.616 was obtained. Using eq. (A1.6), a quality index for the HT hypothetical reference 10×10 cm$^2$ field size, $TPR_{20,10}(10)$ of 0.636 was calculated. This quality index was then used to obtain $k_Q$ = 0.9988 for the A1SL from Table 6.III in TRS 398 [4].

The estimated uncertainty of around 1.5% associated to this approach came from both the estimation of beam quality index (0.3% after [22]) and from the uncertainty in $k_Q$ estimated as 1-1.5% in TRS 398 [4]. In Table A1.2 the different contributions for the uncertainty budget associated with the determination of absolute dose for the *msr* field in HT using an Exradin A1SL chamber are presented.

**Table A1.2** – Uncertainty budget associated with absolute dose determination using an Exradin A1SL chamber with ADCL calibration.

| | **A1SL (%)** | **Uncertainty type** |
|---|---|---|
| Calibration coefficient | 0.7 | B |
| Beam quality correction factor | 0.7 | B |
| Ambient conditions | 0.2 | B |
| Saturation and polarity corrections | 0.2 | B |
| Positioning | 0.2 | B |
| Precision reading | 0.1 | B |
| Machine stability | 0.1 | A |
| Combined expanded uncertainty (k=2) | 2.1 | |

Considering all possible uncertainty sources, a combined expanded uncertainty of 2.1% was obtained (with coverage k=2). The calibration coefficient from the certificate and the beam quality correction factor $k_{Q_{msr}Q_0}^{f_{msr},f_{ref}}$, taken as the weighted mean value from the different approaches,

dominate the uncertainty budget (Table A1.2). The ion recombination correction factor was measured using the two voltages method and it was estimated to be less than 1.002 whereas the polarity effect was negligible for the two A1SL chambers used in this work. Taking into account the alert by Palmans *et al.* [25] that ion recombination can be overestimated in HT, saturation and polarity effects have been included as an uncertainty contribution instead of applied as correction factors. The uncertainty associated to in-water null point positioning of the chamber was reduced from 0.3 mm to 0.1 mm, using the setup megavoltage CT image in HT. Indeed, the MVCT image of the chamber at the water level, observed in all possible views (axial, coronal and sagittal) associated with a proper zooming and sub-millimetric stepwise motion at the console, enabled a reduction of the positioning uncertainty compared to just a visual inspection/assessment of the centering of the chamber at the water level. In both procedures a double check by two medical physicists was performed. Therefore an associated uncertainty of 0.2% in depth dose was estimated. The electrometer reading scale added an estimated uncertainty of 0.1% and the machine stability was estimated to contribute another 0.1% through the relative standard deviation of repetitive measurements in the *msr* HT field.

The *msr* in HT for which reference dosimetry is defined in TRS 483 [1] corresponds to a static delivery. However, Accuray uses as reference a dose plan corresponding to a rotational/helical output to validate the beam modelling in the dedicated TPS. This plan configures a plan class specific reference in the terminology of the Alfonso et al. formalism [3] and TG-148 [6], as already referred in the introduction. It must be stressed that a perfect agreement between helical and static deliveries is assured and maintained, which is also reported in the literature [10,12]. Nevertheless, a factor of 1.003±0.8% is proposed in TG-148 [6] to be used to convert static *msr* field dose into *pcsr* dose. The product of $k_{Q_{msr},Q_0}^{f_{msr},f_{ref}}$ obtained in the present work by this factor of 1.003 equals 0.9975±0.0108, which validates the Accuray practice for absolute dose determination in rotational mode of not correcting the A1SL chamber reading for beam quality.

### A1.3.2 | Relative dosimetry – output factors in HT

For the 18 small fields configured with the binary MLC in HT, the measured FWHM of the longitudinal profiles (Y dimension) differed on average 0.4±0.3 mm from the nominal value. Table A1.3 presents the equivalent clinical square fields ($S_{clin}$) obtained by equating the field dimensions (X and Y) to $\sqrt{XY}$ as recommended in TRS 483 [1] even if for the shadowed cells the ratio Y/X is outside the stated interval [0.7;1.4].

**Table A1.3** – Equivalent square fields calculated by $\sqrt{XY}$, where X is the geometric value of the horizontal upper entry times the scale factor 95/85 and Y is the measured longitudinal FWHM at 10 cm depth of the three field sizes on the vertical left entry. Shadowed cells correspond to $S_{clin}$ where the ratio Y/X is outside the interval [0.7;1.4].

| $S_{clin}$ (cm$^2$) | 1.25 | 2.5 | 3.75 | 5.0 | 7.5 | 10 |
|---|---|---|---|---|---|---|
| **1** | 1.26 | 1.78 | 2.18 | 2.52 | 3.09 | 3.57 |
| **2.5** | 1.98 | 2.80 | 3.43 | 3.96 | 4.85 | 5.60 |
| **5** | 2.82 | 3.98 | 4.88 | 5.63 | 6.90 | 7.96 |

Considering the film output factors as free of corrections, output factor percentage differences from EBT3 film are plotted in Figure A1.1. The output factors for the PTW-60019 microDiamond, the 60018 unshielded diode and the 31016 Pinpoint 3D were calculated accounting for field size corrections interpolated from Table 24 of TRS 483 [1]. The average difference from film, over these

three detectors and all field sizes was 0.3±0.9%. Majorating the uncertainty associated with the reported output correction factors, as recommended by TRS 483 [1] for elongated field sizes (see also the Discussion section), and taking into account the uncertainty associated with film measurements, we can conclude that all measurements are within agreement.

For the A1SL ionization chamber (not contemplated in Table 24 of TRS 483 [1]) the average uncorrected output factor difference from Gafchromic EBT3 film is 0.0±0.7% therefore it can be assumed that the output factors for the A1SL chamber need no correction.

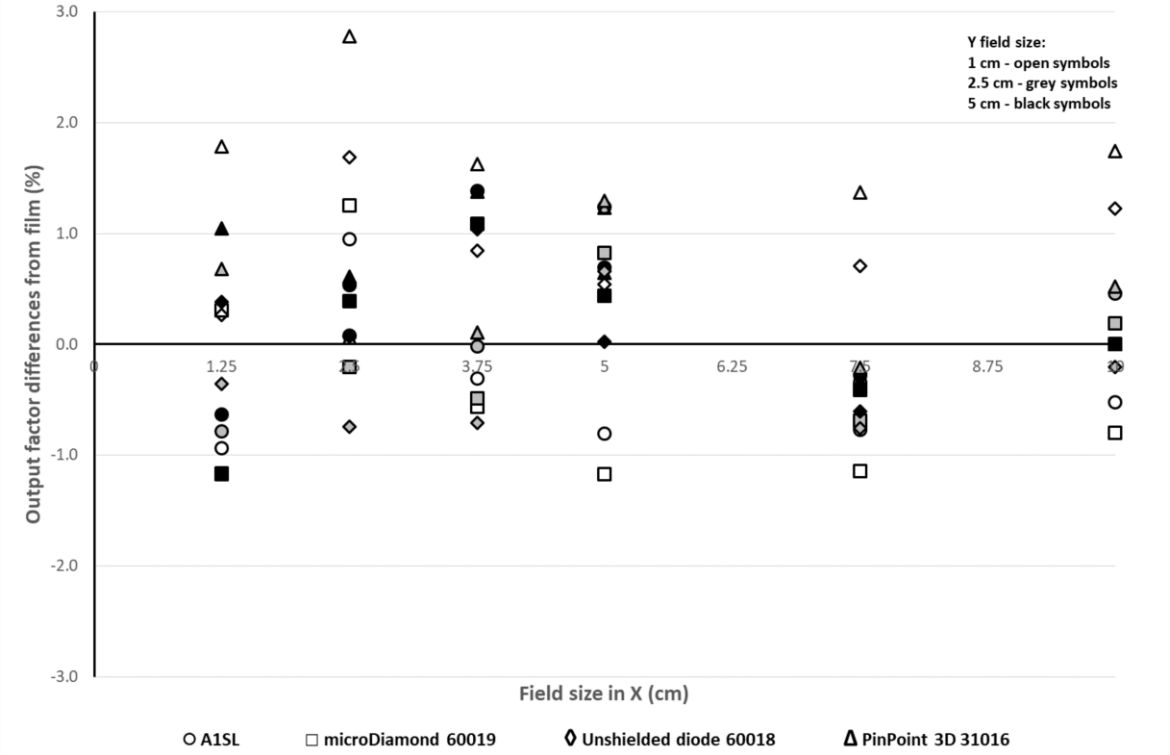


**Figure A1.1** – Output factor percentage differences from EBT3 film, for the 18 configured fields in HT for the Exradin A1SL, the PTW-60019 microDiamond, the 31016 PinPoint 3D, and the 60018 unshielded diode detectors.

## A1.4 | Discussion

### A1.4.1 | Reference dosimetry – msr

Three different approaches have been followed to estimate a beam quality correction factor for the A1SL ionization chamber at *msr* field in HT as this is not reported in TRS 483 [1]. Absolute film dosimetry was the first approach as an alternative from not having a direct calibration coefficient for the chamber at the quality of the *msr* field in HT. The film calibration curve was obtained through the irradiation of film strips with known doses, at reference conditions for a 6 MV photon beam in a Siemens Oncor Avant-Garde linac. This methodology was also followed by Zeverino et al [10]. They determined the beam quality correction factor for different ionization chambers, including the A1SL, under HT msr conditions using EBT2 film. They have irradiated the calibration strips on a Varian 2100C/D linac with a 6 MV, 10×10 $cm^2$ field. In the present work, using the well-established local film dosimetry protocol with multichannel correction the beam quality correction factor for A1SL was obtained. The associated estimated uncertainty accounted for both the experimental and the fit uncertainties, following Devic et al. [17].

The second approach consisted of using two reference chambers – PTW-30013 Farmer and 31013 Semiflex 0.3 cc, for which a generic $k_{Q_{msr}}^{f_{msr},f_{ref}}$ figures in TRS 483 [1], to cross-calibrate the A1SL chamber. The estimated uncertainty associated to this approach accounted for the inherent correlation due to the use of two chambers [4] irradiated in the same setup.

Finally, the concept of a hypothetical reference 10×10 cm$^2$ field size was used to reach a beam quality correction factor for the A1SL chamber in HT. In this approach a $TPR_{20,10}^{HT}$=0.616 was measured as a first step to calculate the quality index of the hypothetical reference field as $TPR_{20,10}(10)$=0.636. Bailat *et al.* [13] reported a value of 0.619 for $TPR_{20,10}^{HT}$ which was converted into $TPR_{20,10}(10)$=0.629, following Jeraj *et al.* [26], which agrees with our derived value within 1.1%. Also Ost *et al.* [11] found a 0.631 value for $TPR_{20,10}(10)$ which is even closer to the one obtained in this work. Strictly speaking, for the application of eq. (A1.5) to FFF beams, $k_{Q,Q_0,}^{f_{ref}}$ should be replaced by the product of two factors. This would account for the difference between the response of the ionization chamber in the hypothetical reference FFF beam of quality $Q^{FFF} = Q$ and that in a beam with flattening filter (WFF) with beam quality $Q^{WFF}$, for which the beam quality index is the same as for the beam quality Q. The main issues that may lead to the differences in $k_{Q,Q_0,}^{f_{ref}}$ data between WFF and FFF beams are differences in water to air stopping-power ratios, perturbation effects and volume averaging due to the difference in charged particle spectra and the non-uniform beam profile of FFF beams. Data in the literature is scarce concerning these effects as noticed in TRS 483 [1]. Given the slow variation of other correction factors with beam quality, the assumption that they are the same for FFF and WFF beams was estimated to result in an additional uncertainty contribution of not more than 0.1% (TRS 483[1]). The uncertainty associated with this approach for the determination of the quality correction factor of the A1SL is sufficiently large to account for this contribution.

Calculating the weighted mean from all the approaches, a value of 0.9945±0.0073 was obtained for the beam quality correction factor of the A1SL chamber agreeing with the mean of $k_{Q_{msr},Q_0}^{f_{msr},f_{ref}}$ values taken from literature, estimated as 0.9937±0.0046 (Table A1.1).

Published values in Table A1.1 for the beam quality correction factor for the A1SL chamber in HT are based in the first approach followed in the present work. In fact, all the referred works used independent methodologies, either experimental (film [10] and alanine [11,12]) or calculated (MC [8,9]). The uncertainty for each of the published correction factors varies from 0.2% for MC calculations to almost 2% for alanine dosimetry.

An unweighted mean over the published values has been calculated. Assuming a rectangular distribution, the estimated associated uncertainty corresponded to L/$\sqrt{3}$ where L is the half-width of the data range. Also in TRS 483 [1] stated uncertainties in published papers were ignored and the assumptions made concerning the estimation of uncertainties are well described in Appendix II. Following once more the rationale of TRS 483 [1], we decided to use the recipe. In fact, computing the weighted mean of the five published values in Table A1.1, would derive an associated uncertainty strongly dominated by the unrealistic small uncertainties reported by MC simulated values. These are just statistical uncertainties, not taking into account other contributing uncertainty factors present in calculations as the influence of transport parameters, geometry description or the MC system used. Concerning the three approaches followed in the present work, we could take control of all estimated uncertainties so the weighted mean and associated global uncertainty was a valid option. Anyway, we could also have included our results into the data range with the published ones. Taking the average of all the nine values in Table A1.1, and considering again a rectangular distribution, which attributes to each of the values the same probability of being the

"correct" beam quality correction factor for the A1SL chamber, we would obtain 0.9943±0.0051. This could also be considered the global result of the present work.

### A1.4.2 | Relative dosimetry – output factors in HT

A set of different detectors has been used in this work, following the TRS 483[1] recommendation of using different types of detectors so that redundancy in the results can provide more confidence in performing small field dosimetry.

In this context, an exercise was carried out in order to explore and compare small fields output factors in HT which had the ultimate aim of reassuring the correct progress from reference dosimetry to relative dosimetry with the A1SL chamber in HT.

Concerning the calculation of equivalent small clinical field sizes, the way Y and X field sizes were determined should be justified. Due to the effect of penumbra overlap in conventional linear accelerators, for field sizes smaller than the charged particle lateral diffusion distance, the measured FWHM overestimates the field size [1]. For our HT, we have estimated the lateral charged particle equilibrium range as 9.4 mm. So the profile widths would considerably differ from geometrical field sizes for field sizes smaller than 1.9 cm. From the measured FWHM of the Y profiles, where the smaller geometrical field size is 1.1 cm at the measuring depth, we have measured FWHMs that differed less than 0.2 mm from the nominal value which may come from the different collimation geometry in HT. For the smallest X profile, the nominal value at the measuring depth was 1.4 cm corresponding to 2 open leaves. As the binary MLC in HT is below the jaws, the effect of penumbra overlap will be lower. These two facts (larger field size in X and MLC positioning along the beam line) justifies the argument of using just geometry to define the width of the X profile.

Taking into account the considerations from Appendix II of TRS 483 [1], the tabulated values of the field size dependent output correction factors are weighted mean values from MC simulations and experimental data, for linac 6 MV photon beams. Combined expanded uncertainties with coverage factor k=2 have been considered. Global values of relative standard uncertainties for detector groups are also published in Table 37 of that appendix, for the same range of equivalent square field sizes, 0.4 to 8 cm, from where the interpolated values used in this work have been obtained. For HT machines the same fitting parameters for 6 MV beams have been considered in TRS 483 [1] with proper normalization to the *msr* field. For the PTW-60019 microDiamond and the 60018 unshielded diode, the uncertainties range from 0.9% at 0.4 cm field size to 0.3% at 5 cm field size. For the PTW-31016 PinPoint chamber, included in the micro IC group, the uncertainties range from 3.2% at 0.5 cm field size to 0.4% at 5 cm field size. Having in mind that most of the equivalent square small fields configured in HT for this work do not respect the field size interval for the $S_{clin}$ calculation, a higher uncertainty associated to the output correction factor was considered, as recommended.

Considering the film output factors as free of corrections, the field size dependent correction factors for the Exradin A1SL have been calculated. Taking into account both the film uncertainty and the measurement influencing quantities, the output correction factors for the A1SL chamber for the considered field size range are compatible with unity. This means that the ratios of A1SL chamber readings do not need further correction in order to be taken as output factors for the considered small clinical field sizes when normalized to the *msr* field in HT.

## A1.5 | Conclusions

Exradin A1SL is the standard reference chamber included in the HT dosimetry package. Facing the lack in TRS 483 of a generic beam quality correction factor to account for quality and geometry of the *msr* field in HT for this ionization chamber, different approaches were followed tracking the concepts and recommendations of TRS 483 through: i) independent film dosimetry; ii) cross-calibration with reference chambers; iii) $TPR_{20,10}(10)$ of the hypothetical 10×10 cm$^2$ reference field corresponding to the HT beam quality and using the TRS 398 $k_Q$ factor. All three approaches led to a weighted mean value for the beam quality correction factor for reference dosimetry in HT with A1SL chamber of 0.9945±0.0073, agreeing with other averaged value from published results based on Monte Carlo simulations, alanine or film independent dosimetry. A combined expanded uncertainty of 2.1% with coverage factor of k=2 was derived for reference static dosimetry with the A1SL chamber in the HT machine. These results validate the Accuray practice of not correcting the A1SL chamber reading for quality when performing absolute dose determination of the reference calibration plan in rotational mode. In fact, the derived $k_{Q_{msr}Q_0}^{f_{msr},f_{ref}}$ multiplied by a conversion factor from static to rotational references is compatible with unity within 1%.

For relative dosimetry in HT machine, eighteen small fields were configured. Output factors for these small fields, having $f_{msr}$ as reference, were measured for a PTW-31016 PinPoint 3D, a 60018 unshielded diode, a 60019 microDiamond, a Exradin A1SL chamber, and Gafchromic EBT3 film. Field size dependent output correction factors for the calculated equivalent square field sizes were interpolated from published values in TRS 483. Given the level of uncertainty, all measurements were within agreement. Considering film output factors as free of corrections, the ratio of the A1SL chamber measurements for the small clinical field sizes relative to the *msr* field in HT can be taken as output factors, needing no further corrections.

## A1.6 | Acknowledgments

Part of this work was supported by FCT, the Portuguese Foundation for Science and Technology, through a PhD scholarship, reference SFRH/BD/118929/2016 and the project POCI-01-0145-FEDER-028030.

The authors wish to thank Pedro Andreo his expert advice on the useful discussion about estimating uncertainties.

## A1.7 | Conflict of interest

The authors have no conflict to disclose.

## A1.8 | References

1. Alfonso R, Andreo P, Capote R, et al., International Atomic Energy Agency (IAEA). Dosimetry of Small Static Fields Used in External Beam Radiotherapy: An International Code of Practice for Reference and Relative Dose Determination. IAEA TRS-483. Vienna: International Atomic Energy Agency; 2017
2. Palmans H, Andreo P, Huq MS, et al. Dosimetry of small static fields used in external photon beam radiotherapy: Summary of TRS-483, the IAEA–AAPM international Code of Practice for reference and relative dose determination. Med Phys. 2018;45(11):e1123-e1145. doi:10.1002/mp.13208

3. Alfonso R, Andreo P, Capote R, et al. A new formalism for reference dosimetry of small and nonstandard fields. Med Phys. 2008;35(11):5179-5186. doi:10.1118/1.3005481
4. International Atomic Energy Agency (IAEA). Absorbed Dose Determination in External Beam Radiotherapy: An International Code of Practice for Dosimetry Based on Standards of Absorbed Dose to Water. IAEA TRS-398. Vienna: International Atomic Energy Agency; 2000
5. Almond PR, Biggs PJ, Coursey BM, et al. AAPM's TG-51 protocol for clinical reference dosimetry of high-energy photon and electron beams. Med Phys. 1999;26(9):1847-1870. doi:10.1118/1.598691
6. Langen KM, Papanikolaou N, Balog J, et al. QA for helical tomotherapy: Report of the AAPM Task Group 148a). Med Phys. 2010;37(9):4817-4853. doi:10.1118/1.3462971
7. Roy M Le, de Carlan L, Delaunay F, et al. Assessment of small volume ionization chambers as reference dosimeters in high-energy photon beams. Phys Med Biol. 2011;56(17):5637-5650. doi:10.1088/0031-9155/56/17/011
8. Thomas SD, Mackenzie M, Rogers DWO, Fallone BG. A Monte Carlo derived TG-51 equivalent calibration for helical tomotherapy. Med Phys. 2005;32(5):1346-1353. doi:10.1118/1.1897084
9. Sterpin E, Mackie TR, Vynckier S. Monte Carlo computed machine-specific correction factors for reference dosimetry of TomoTherapy static beam for several ion chambers. Med Phys. 2012;39(7Part1):4066-4072. doi:10.1118/1.4722752
10. Zeverino M, Agostinelli S, Pupillo F, Taccini G. Determination of the correction factors for different ionization chambers used for the calibration of the helical tomotherapy static beam. Radiother Oncol. 2011;100(3):424-428. doi:10.1016/j.radonc.2011.08.044
11. De Ost B, Schaeken B, Vynckier S, Sterpin E, den Weyngaert D. Reference dosimetry for helical tomotherapy: Practical implementation and a multicenter validation. Med Phys. 2011;38(11):6020-6026. doi:10.1118/1.3651496
12. Gago-Arias A, Rodríguez-Romero R, Sánchez-Rubio P, et al. Correction factors for A1SL ionization chamber dosimetry in TomoTherapy: Machine-specific, plan-class, and clinical fields. Med Phys. 2012;39(4):1964-1970. doi:10.1118/1.3692181
13. Bailat CJ, Buchillier T, Pachoud M, Moeckli R, Bochud FO. An absolute dose determination of helical tomotherapy accelerator, TomoTherapy High-Art II. Med Phys. 2009;36(9Part1):3891-3896. doi:10.1118/1.3176951
14. Ferreira BC, Lopes MC, Capela M. Evaluation of an Epson flatbed scanner to read Gafchromic EBT films for radiation dosimetry. Phys Med Biol. 2009;54(4):1073-1085. doi:10.1088/0031-9155/54/4/017
15. Micke A, Lewis DF, Yu X. Multichannel film dosimetry with nonuniformity correction. Med Phys. 2011;38(5):2523-2534. doi:10.1118/1.3576105
16. Mayer RR, Ma F, Chen Y, et al. Enhanced dosimetry procedures and assessment for EBT2 radiochromic film. Med Phys. 2012;39(4):2147-2155. doi:10.1118/1.3694100
17. Devic S, Tomic N, Lewis D. Reference radiochromic film dosimetry: Review of technical aspects. Phys Med. 2016;32(4):541-556. doi:10.1016/j.ejmp.2016.02.008
18. Dreindl R, Georg D, Stock M. Radiochromic film dosimetry: considerations on precision and accuracy for EBT2 and EBT3 type films. Z Med Phys. 2014;24(2):153-163. doi:10.1016/j.zemedi.2013.08.002
19. Casanova Borca V, Pasquino M, Russo G, et al. Dosimetric characterization and use of GAFCHROMIC EBT3 film for IMRT dose verification. J Appl Clin Med Phys 2013;14:4111. doi:10.1120/jacmp.v14i2.4111
20. Khachonkham S, Dreindl R, Heilemann G, et al. Characteristic of EBT-XD and EBT3 radiochromic film dosimetry for photon and proton beams. Phys Med Biol 2018;63:65007. doi:10.1088/1361-6560/aab1ee
21. Thomas SJ, Aspradakis MM, Byrne JP, et al. Reference dosimetry on TomoTherapy: an addendum to the 1990 UK MV dosimetry code of practice. Phys Med Biol. 2014;59(6):1339-1352. doi:10.1088/0031-9155/59/6/1339
22. Palmans H. Determination of the beam quality index of high-energy photon beams under nonstandard reference conditions. Med Phys. 2012;39(9):5513-5519. doi:10.1118/1.4745565
23. Georg D, Knoos T and Mcclean B. Current status and future perspective of flattening filter free photon beams. Med Phys. 2011;38(3):1280-1293. doi: 10.1118/1.3554643

24. Fogliata A, Fleckenstein J, Schneider F, et al. Flattening filter free beams from TrueBeam and Versa HD units: Evaluation of the parameters for quality assurance. Med Phys. 2016;43(1):205-212. doi:10.1118/1.4938060

25. Palmans H, Thomas RAS, Duane S, Sterpin E, Vynckier S. Ion recombination for ionization chamber dosimetry in a helical tomotherapy unit. Med Phys. 2010;37(6):2876-2889. doi:10.1118/1.3427411

26. Jeraj R, Mackie TR, Balog J, Olivera G. Dose calibration of nonconventional treatment systems applied to helical tomotherapy. Med Phys. 2005;32(2):570-577. doi:10.1118/1.185501526

# APPENDIX II

# Independent verification of the pre-installed beam model in helical tomotherapy

Accepted for poster presentation at the International Conference on Advances in Radiation Oncology (ICARO-3), 16-19 February 2021 (virtual event)

MC Lopes[a], T Santos[a,b], T Ventura[a], M Capela[a]

[a] Medical Physics Department, IPOCFG, E.P.E., Coimbra, Portugal
[b] Physics Department, University of Coimbra, Coimbra, Portugal

## A2.1 | Background and Objective

Helical Tomotherapy (HT) is a pre-commissioned equipment. The pre-installed beam model only needs to be verified by the customer as part of commissioning process. With the aim of supporting medical physicists in this task, the AAPM Task Group 148 report was published in 2010 [1]. The joint IAEA and AAPM international code of practice (CoP) for small static fields dosimetry – TRS 483 [2] – builds on the established reference dosimetry and extends it to non-reference fields by introducing the concept of machine specific reference (*msr*) field.

To report the local procedures for independent verification of the pre-installed beam model – gold standard (GS) dosimetry in HT – was the aim of the present work.

## A2.2 | Methods

The standard HT quality assurance package includes: 2 ionization chambers Exradin A1SL, one TomoElectrometer with 8 channels and a 2D TomoScanner water tank, all from Standard Imaging. The longitudinal and transverse dose profiles as well as the percent depth doses for the three field sizes – 1, 2.5 and 5 cm length – are measured by Accuray during the installation process with the 2D TomoScanner water tank, and compared to the GS dosimetry corresponding to the pre-installed beam model. Beam adjustment is assessed through 1D gamma-function analysis where the Accuray tolerances are set at: i) 2% Dose Difference/1% Distance-To-Agreement (which converts to 0.25 mm for the 2.5 cm field size profile at isocentre, for instance) for longitudinal profiles at 15 mm depth; ii) 2%/1 mm for transverse profiles also at 15 mm depth and iii) PDD analysis is based on dose ratio with tolerance of 2% from 10 to 200 mm depth. Full width at half maximum (FWHM) of longitudinal profiles should also be within 1% of the correspondent length in GS. For transverse profiles, it is the full width at 25% maximum (FWQM), also at 15 mm depth, that should be within 1% of FWQM of the corresponding GS profile.

The local dosimetry package included a MP3-T motorized water phantom from PTW, with the aim of being independent from Accuray procedures. Using MP3-T with the beam data acquisition software MEPHYSTO mcc 3.2 and the standard A1SL ionization chamber, dose profiles measured at installation have been repeated and compared with GS curves. To further enhance independency the measurement of the same profiles was repeated using the PTW 31016 PinPoint 3D chamber in orthogonal alignment and using the standard PTW detector holder for the MP3-T phantom.

For dose calibration the *msr* concept has been used and an independent dosimetry audit was carried out by an IAEA expert.

## A2.3 | Results and Discussion

PDD results for the two data sets – using A1SL and PTW 31016 PinPoint 3D chambers with the PTW MP3-T motorized water phantom – were within 1% to the GS using the 2D TomoScanner, well below the required Accuray ratio criteria of 2% PDD ratios.

Longitudinal profiles for the three field sizes – 1, 2.5 and 5 cm – using both chambers complied almost everywhere with the 1%/1 mm distance-to-agreement gamma criteria when GS is taken as reference, except in a limited region outside the larger field, where the standard 2% dose criterion was required.

Transverse profiles exhibiting the FFF characteristic cone-shape were measured with all MLC leaves open which corresponds to 40 cm width at isocentre. FWQM of these profiles at 10 cm depth

met 1% tolerance from GS for both measurement sets, even at a deeper depth than 15 mm as required by acceptance testing procedure (ATP) for this parameter.

For absolute dose calibration the percent deviation from the external audit dosimetry system was 0.2%.

## A2.4 | Conclusions

The installation of a helical Tomotherapy machine and the publication of the CoP TRS 483 motivated the presented work. Independent verification of the pre-installed beam model in HT was reported and proven to be in closer agreement than the Accuray acceptance tolerance levels. The independency from Accuray procedures was based both in a different water phantom, the PTW MP3-T with standard ionization chambers (Exradin A1SL) and using non-standard detectors like PTW 31016 PinPoint 3D.

This is the first dosimetry results presented for HT using the MP3-T, although other phantoms have also been tested [3].

## A2.5 | References

[1] LANGEN, K.M., et al., QA for helical tomotherapy: Report of the AAPM Task Group 148, Med. Phys. 37 (2010) 4817–4853

[2] INTERNATIONAL ATOMIC ENERGY AGENCY, Dosimetry of small static fields used in external beam radiotherapy, Technical Reports Series No. 483, IAEA, Vienna (2017)

[3] PENG J. L., ASHENAFI M. S., MCDONALD D. G., VANEK K. N., Assessment of a three dimensional (3D) water scanning system for beam commissioning and measurements on a helical tomotherapy unit, J. Appl. Clin. Med. Phys. 16 (2015) 51-68