Physics and Imaging in Radiation Oncology 32 (2024) 100666

Contents lists available at ScienceDirect

Physics and Imaging in Radiation Oncology

journal homepage: www.sciencedirect.com/journal/physics-and-imaging-in-radiation-oncology

Original Research Article

A commissioning protocol for portal imaging-based radiotherapy in vivo dosimetry systems

Marco Esposito [a,b], Riccardo Baldoni [c], Evy Bossuyt [d], Sara Bresciani [e], Catharine H Clark [f,g], Matthew Jones [h], Stephen Kry [i], Joseph Perry [h], Jeroen van de Kamer [j], Dirk Verellen [d,k], Nuria Jornet [l,*]

[a] Medical Physic, The Abdus Salam International Centre for Theoretical Physics, Trieste, Italy
[b] S.O.C. Fisica Sanitaria- Azienda Sanitaria USL Toscana Centro,Firenze, Italy
[c] U.OC-Fisica Medica e alte tecnologie, AST Pesaro Urbino, Pesaro, Italy
[d] Medical Physics, Iridium Netwerk, Antwerp, Belgium
[e] Medical Physics, Candiolo Cancer Institute—FPO IRCCS, Turin, Italy
[f] University College London, London, UK
[g] National Physical Laboratory, Teddington, UK
[h] Royal Surrey NHS Foundation Trust,Guildford, UK
[i] The University of Texas M.D. Anderson Cancer Center, Houston, TX, USA
[j] The Netherlands Cancer Institute, Amsterdam, the Netherlands
[k] Antwerp University, Faculty of Medicine and Health Sciences, Antwerp, Belgium
[l] Servei de Radiofisica i Radioprotecció, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain

ARTICLE INFO

Keywords:
"In vivo" dose measurements
EPID
QA
Dosimetry

ABSTRACT

Background and Purpose: With the availability of commercial electronic portal imaging detector-based in vivo dosimetry (EPID-based IVD) solutions, many radiotherapy departments are adopting this technology. However, comprehensive commissioning guidance is lacking. This study aims to provide a protocol for testing the accuracy and sensitivity of EPID-based IVD systems.

Material and methods: The protocol was tested across four institutions using two different systems. Accuracy was evaluated with homogeneous slab phantoms using different square regular fields, and clinical plans in a CIRS lung phantom. Multiple forward and back-projected algorithm implementations were assessed for different energies. Sensitivity analysis in the lung phantom examined responses to setup errors, anatomical variations, and delivery errors.

Results: In homogeneous phantoms, over 85 % of pixels passed the 5 %/2mm gamma criteria, except for the 2x2 cm^2 field. In the lung phantom, all systems and implementations achieved over 95 %-pixel pass rates at the 2 %/2mm criterion for volumetric modulated arc therapy (VMAT) plans. For conformal radiation therapy (3DCRT) plans, one system implementation showed poor accuracy, with over 90 % agreement only at the 5 %/2mm criterion. Considering all systems and implementations, average sensitivity and specificity for CRT plans ranged from 0.92 and 0.42 (at 2 %/2mm) to 0.71 and 0.52 (at 5 %/2mm), while for VMAT plans ranged from 0.41 and 0.81 (at 2 %2mm) to 0.37 and 0.81 (at 5 %/2mm).

Conclusion: We successfully developed a protocol to commission EPID IDV systems. It was found that not all systems and implementations achieved satisfactory accuracy and sensitivity, emphasising the need for thorough commissioning and benchmarking.

1. Introduction

In vivo dosimetry (IVD) assesses the agreement between the planned dose and that delivered to the patient during radiotherapy treatment. For several decades, international organizations have recommended its use [1–5], and national and international regulators are starting to

* Corresponding author at: Servei de Radiofísica i Radioprotecció, Hospital de la Santa Creu i Sant Pau, Barcelona, Spain.
E-mail address: njornet@santpau.cat (N. Jornet).

https://doi.org/10.1016/j.phro.2024.100666
Received 13 May 2024; Received in revised form 27 October 2024; Accepted 28 October 2024
Available online 1 November 2024

require it [2].

Electronic Portal Imaging Device (EPID) based IVD is an ensemble of computational techniques that, using the signal collected by the portal imager after passing through the patient (transit dosimetry), are used to detect clinically relevant errors before the therapy effectiveness is compromised [5–7]. It can be divided into two classes: forward-projection (FP) and back-projection (BP) methods. In the first class, the signal measured by the EPID is converted to dose, and compared to a prediction by the FP algorithm. The comparison is usually based on 2D Gamma Agreement Index (GAI), but profile analysis and point dose differences can also be used. In the second class, the BP methods reconstruct the absorbed dose in the patient anatomy, by back-projecting the signal measured by the EPID to either a point, a plane or in 3D. BP reconstructed dose can be compared directly with the planned dose using point dose difference or 2D, 3D gamma agreement or Dose Volume Histogram (DVH) difference.

EPID IVD has shown its potential to detect errors in single and multi-institutional studies [8–15] and for remote auditing [16]. In the last ten years the number of commercial solutions available on the market as well as the number of publications about the EPID IVD implementation have increased, signalling an increasing interest of the radiotherapy community on this topic. However, despite the wide availability of both the EPID and EPID based IVD systems, the broad clinical application of this methodology is still limited to few centres with considerable experience in this area. One reason for the difficulties in implementing an EPID IVD clinical program is the lack of guidelines for acceptance and commissioning of these systems. Over the past several years, patient specific quality assurance devices have come under increasing scrutiny for being unable to detect substantial dose deviations that manifest as important dose errors [17]. In addition to identifying a plan delivery as acceptable when it is not, EPID IVD programs, because of the additional processing steps, are also known to be at risk of flagging a dose delivery session as unacceptable when it is acceptable [6]. The evaluation of the IVD algorithm accuracy is a crucial step before its clinical implementation. Ideally, the tolerance levels for error determination should be set to identify clinically relevant failures, and to avoid the false positives and false negatives arising from EPID IVD algorithm failure, or from small inaccuracies due to multiple sources, without a clear clinical impact. The need to set correctly these tolerances is critical to ensure safe patient treatments while also keeping efficiency for routine clinical use.

A procedure for commissioning and performance testing for EPID based IVD software has been developed in this work. This procedure utilizes commonly available phantoms and simple tests to assess both the accuracy and the sensitivity of the algorithms. It is designed to be applicable to any class of EPID IVD algorithm. The goal of this work was to provide clinical medical physicists, developers, and researchers with a standardized and practical framework for testing and optimising any implementation of EPID based IVD systems.

2. Materials and methods

This work has been conducted by an ESTRO working group, created after the 2018 ESTRO Physics workshop in Malaga, for critically evaluating EPID IVD implementation. Four centres from three different countries in Europe participated in this study, the characteristics of the EPID dosimetry systems, linear accelerator (linac), treatment planning systems (TPS) and phantoms used are reported in Table 1. The EPID IVD solutions used in this study have been described extensively elsewhere [18–20]. The measurement protocol was divided in two parts. The first part focused on testing the accuracy of the IVD system and its implementation across a range of planning techniques, from basic to complex, using different phantoms. This included a set of square irradiation fields in homogeneous water-equivalent slab phantoms, SLABthick (30 cm thick) and SLABthin (10 cm thick), as well as in an anthropomorphic lung phantom, ANTHROinhom (CIRS lung phantom). Additionally, conformal radiotherapy (3DCRT), intensity modulated radiotherapy (IMRT) or volumetric modulated arc therapy (VMAT) clinical plans were tested in the anthropomorphic phantom. If the anthropomorphic phantom was unavailable, SLABthick was used as an alternative. The second part of the study examined the overall sensitivity of the IVD system in detecting various errors, such as anatomical changes, setup errors and linac delivery errors, by applying a set of clinical plans to ANTHROinhom phantom.

2.1. Treatment planning

All phantoms underwent a simulation CT using the institution's clinical scanning protocol. For the slab phantoms, the density was adjusted to match their nominal values.

The following plans were created by each institution, using a 2x2x2 mm^3 dose calculation grid. To begin with, a set of square irradiation fields of different sizes (20x20 cm2, 10x10 cm2, 5x5cm2, and 2x2cm2) were created for the SLABthick and SLABthin phantom. For these fields, the gantry was set to 0° and the monitor units (MU) were fixed at 100, with the plan isocenter located at the center of the phantom. Subsequently, clinical plans were generated using ANTHROinhom phantom. Participants contoured a 2 cm radius spherical target and expanded it isotropically by 5 mm to form the planning target volume (PTV). The first clinical plan involved a 4-field box technique to treat the PTV, prescribing a dose of 2 Gy at the isocenter placed at the PTV center. The beams were labelled as follows: H1 (Gantry 0°), H2 (Gantry 180°), H3 (Gantry 90°), and H4 (Gantry 270°).

The second clinical plan was either an IMRT or VMAT plan, using the same lung target as in the 3DCRT plan, and adhering to institutional preferences and guidelines. The plan isocenter was placed at the PTV center, with a dose prescription of 2 Gy per fraction covering 95 % of the PTV.

Afterwards, various errors were introduced into each 3DCRT and IMRT/VMAT plan to assess their impact. These errors included: anatomic variations, where changes in patient contour were made by adding 1 cm, 2 cm, and 3 cm boluses (20x20 cm^2) to the phantom

Table 1
Summary of EPID based IVD algorithms, software, TPS, dose computation algorithms, CT scanning protocols, linacs and EPID conficurations used by the participants.

Centre	Software	Algorithm	Beam Energy	TPS	Dose computation algorithm	CT scanning parameters	Linac	EPID
A	PerFRACTION	FP	6 MV 15 MV	Eclipse	AAA	120 kV, 100mAs, 2 mm	Varian Truebeam	aS1200
B	PerFRACTION	FP	6 MV	Eclipse	AAA	120 kV, 100mAs, 2 mm	Varian Truebeam	aS1000
C	PerFRACTION	FP	6 MV 10 MV 15 MV	Eclipse	AAA	120 kV, 34mAs, 2 mm	Varian Truebeam	aS1000
D	Dosimetry Check	BP	6 MV 10 MV	Monaco	Monte Carlo, dose to medium	120 kV, 100 mAs, 2.5 mm	Elekta Synergy	Iview-GT aSi

Note: FP and BP refer to forward projection and back projection algorithms.

surface, covering the ipsilateral lung. Setup errors: lateral shifts of the treatment table from the isocenter by 5 mm, 10 mm, and 15 mm. Linac delivery errors: collimator rotations of 5°, 10°, and 15°. Plans with introduced errors were evaluated by comparing their dose distributions with the original plans to identify discrepancies with potential clinical impact [21]. This study focused on dose variations within the PTV (ΔD50% ΔD98% and ΔD2%). According to ICRU 83, a dose variation greater than 3.5 % in the D50% was considered clinically significant, potentially leading to clinically relevant effects [21,22]. In the absence of established guidelines for ΔD98% and ΔD2%, a threshold double that for ΔD50% from ICRU 83 was adopted: ±7%. Thus, an error was deemed clinically significant if it resulted in any of the following dose variations within the PTV:

$$|\Delta D|_{50\%} > 3.5\ \%;\ \text{or}\ |\Delta D_{98\%}| > 7\%;\ \text{or}\ |\Delta D_{2\%}| > 7\ \%\ (1)$$

In this work, the sensitivity of the EPID algorithm to identify the clinically significant errors was tested; otherwise, errors resulting in dose variation below these specified thresholds in (1) were not assessed.

2.2. Measurements

Before irradiating the phantoms, the linac was recalibrated under reference conditions. The phantoms were positioned at the linac isocenter as per the treatment plans. Before delivering treatment, we verified phantom positioning using pre-treatment imaging (preferably cone beam computed tomography) and applied setup corrections if necessary.

2.3. Accuracy analysis

The accuracy of the EPID algorithms was assessed with the plans described previously. For the forward-projection algorithm the dose associated with the EPID was compared with the system prediction; for the back-projection algorithm the 2D TPS dose on the coronal plan passing through isocentre was compared with the reconstructed dose inside the phantom at the same plane. To check the accuracy of the EPID transit algorithms the agreement between measured and computed dose distributions was evaluated with the 2D Gamma function using the following criteria: (2 %/2mm, 3 %/2mm, 5 %/2mm) normalized to the maximum, a threshold of 30 % of the maximum dose was applied. Three gamma agreement index (GAI) thresholds were considered: >95 %, >90 %, >85 %.

2.4. Sensitivity analysis

For the Sensitivity analysis, the threshold for error determination was set to 90 % of pixels passing for all 2D gamma criteria (2 %/2mm, 3 %/2mm, 5 %/2mm), a threshold of 30 % of the maximum dose was applied. Based on this criteria and tolerance, a true positive outcome (TP) was one where there was a clinically significant error in the TPS simulation and < 90 % of pixels passing in the EPID measurement. A false positive error (FP) was a case with a non-clinically significant error in the TPS simulation but < 90 % of pixels passing the EPID measurement. A false negative (FN) was a case with a clinically significant error in the TPS simulation but > 90 % of pixels passing the EPID measurement. Finally, a true negative (TN) was a case with non-significant error in the TPS simulation and had > 90 % of pixels passing in the EPID measurement. Sensitivity and specificity have been computed using the following formulas: Sens = TP/(TP + FN); Spec: TN/(TN + FP).

3. Results

3.1. Accuracy

As shown in Table 2, among all the IVD system/implementations, the smallest field (2x2) exhibited the least accurate results, with only 5/9 of the models achieving > 85 % of pixels passing for the least stringent 5 %/2mm criterion. For all other fields, all the systems and implementations fulfilled at least > 85 % of pixels passing the least stringent criterion. The forward projection algorithm showed better accuracy in the homogeneous phantoms: more than 70 % of tested fields showed > 95 %

Table 2
Gamma Agreement Index (GAI) for various field sizes (2x2 cm^2, 5x5 cm^2, 10x10 cm^2, and 20x20 cm^2) across different centers for the two homogeneous water equivalent phantoms SLABthin (first line) and SLABthick (second line bold) phantom. Note that Centre D tested two different algorithm implementations for the 6 MV beam. In red values < 90 %.

Field Size		2x2 cm			5x5 cm			10x10 cm			20x20 cm		
GAI criteria		2% 2mm	3% 2mm	5% 2mm	2% 2mm	3% 2mm	5% 2mm	2% 2mm	3% 2mm	5% 2mm	2% 2mm	3% 2mm	5% 2mm
Centre A	FP	100	100	100	89	100	100	86	99	100	53	98	100
	6 MV	**100**	**100**	**100**	**97**	**100**	**100**	**99**	**100**	**100**	**100**	**100**	**100**
	FP	63	63	66	77	87	97	83	93	97	90	98	99
	15 MV	**63**	**63**	**69**	**93**	**98**	**99**	**97**	**99**	**100**	**99**	**100**	**100**
Centre B	FP	100	100	100	100	100	100	100	100	100	86	87	99
	6 MV	**90**	**94**	**100**	**89**	**96**	**100**	**60**	**98**	**100**	**67**	**98**	**99**
Centre C	FP	93	98	100	94	99	100	97	100	100	100	100	100
	6 MV	**89**	**92**	**100**	**89**	**94**	**100**	**85**	**99**	**100**	**100**	**100**	**100**
	FP	94	94	100	93	95	100	99	100	100	99	99	100
	10 MV	**92**	**94**	**100**	**95**	**98**	**100**	**100**	**100**	**100**	**100**	**100**	**100**
	FP	80	84	89	94	99	100	99	100	100	100	100	100
	15 MV	**70**	**77**	**83**	**99**	**100**	**100**	**98**	**100**	**100**	**100**	**100**	**100**
Centre D	BP1	60	60	60	74	79	88	78	82	87	66	80	86
	6 MV	**33**	**97**	**51**	**76**	**83**	**92**	**81**	**91**	**95**	**19**	**63**	**99**
	BP2	78	83	87	84	90	100	77	83	90	50	83	93
	6 MV	**100**	**100**	**100**	**100**	**100**	**100**	**79**	**87**	**97**	**32**	**57**	**99**
	BP	67	67	67	82	85	92	90	94	97	92	95	98
	10 MV	**50**	**50**	**56**	**76**	**77**	**89**	**78**	**87**	**94**	**47**	**62**	**85**

of pixels passing with the most stringent 2 %/2mm criterion as compared to only 23 % of the back-projection algorithms.

In presence of inhomogeneities, for the 3DCRT plans, Centre A with a FP algorithm and Centre D with a BP algorithm, showed high accuracy, with > 95 % of points passing the most stringent gamma criterion at 2 %/2mm (Fig. 1). The Centre B, otherwise, produced less accurate results with a FP algorithm: only the 5 %/2mm criterion showed > 90 % of pixels in agreement. With the VMAT plans, all the algorithms tested by the three Centres have > 95 % of pixels in agreement even with the most stringent 2 %/2mm criteria.

3.2. Sensitivity

As depicted in Tables 3 and 4, EPID algorithms, both forward and back-projections, were more sensitive to errors in 3DCRT plans (average sensitivity between 0.92 at 2 %/2mm and 0.71 at 5 %/2mm, compared to VMAT plans (average sensitivity between 0.41 at 2 %/2mm and 0.37 at 5 %/2mm). Conversely, VMAT plans exhibited better specificity (average 0.81), than 3DCRT plans (average specificity between 0.42 at 2 %/2mm and 0.52 at 5 %/2mm) The lower accuracy observed at Centre C in the 3DCRT lung plans is evident in the sensitivity and specificity evaluations, where all introduced errors (both relevant and irrelevant) resulted in less than 90 %-pixel agreement, yielding a Sens and Spec score of 1.0 (Fig. 2). The back-projection algorithm demonstrated slightly higher sensitivity to error detection: with the 2 %/2mm metric, the sensitivity to detect significant errors both in 3DCRT and VMAT plan was above 0.8 for the back-projection algorithm, whereas for the forward projection algorithm, sensitivity remained high for 3DCRT plans but fell below 0.5 for VMAT plans.

4. Discussion

In this study, we proposed and validated a protocol for evaluating EPID IVD systems, applicable to both FP and BP algorithms. Although not exhaustive, it included tests of increasing complexity for basic functionality and sensitivity analysis. This protocol balanced test completeness with resource constraints, requiring about two hours of linac irradiation time, making it feasible for most clinics.

Recently, the AAPM task group 307 [5] recommended validating EPID dosimetry software before clinical use, focusing on both accuracy and sensitivity. However, they did not specify the tests to perform.

Previous studies on sensitivity and specificity of different EPID transit dosimetry systems using phantoms have employed different methodologies and forced different types of delivery errors, complicating direct comparisons [18,19,23–28]. Generally, all studies found that sensitivity and specificity depend on the clinical scenario (anatomic site), error magnitude, and comparison metric. Bedford et al. [23] found similar accuracy between FP and BP methods but noted that the FP method was more sensitive to shift errors. Conversely, Mijnheer et al. [24] and Yedekci et al. [25] reported low sensitivity to setup and shift errors for BP methods.

Our protocol addresses these gaps by focusing on prevalent challenges associated with EPID-based IVD, such as low accuracy, oversensitivity to minor clinically relevant errors, and poor sensitivity to setup and shift errors [5–8]. This commissioning approach aims to enhance dosimetry system reliability, potentially guiding industry testing and optimization before market launch.

The proposed sensitivity analyses were applied specifically to lung treatments and should be considered valid only for this clinical scenario. While the proposed validation methodology can be extended to other contexts, it is crucial to use appropriate anthropomorphic phantoms that accurately replicate the specificities of the anatomic site of interest.

The study also aimed to conduct a pilot test to quickly assess the feasibility of implementing the protocol in an international multicenter setting. The four centers included were selected for their extensive experience in the clinical use of EPID IVD and their ability to complete all the evaluation tests within three months. While our results are limited by the small number of centers included in the study, and excluded FFF beam, they reveal significant differences that underscore the need for rigorous commissioning before clinical implementation. The proposed protocol has demonstrated its ability to identify these variations, validating its effectiveness. These findings justify applying the proposed protocol to a larger multicenter study. Currently, the Italian Association for Medical Physics working group is conducting a multicenter study using this protocol to assess differences between different EPID based IVD systems and implementations. Results from this broader study are anticipated by the end of next year, which will provide further insights and reinforce the need for comprehensive commissioning practices.

An aspect not directly addressed in this study is the optimization of the comparison metric for error detection. We compared the sensitivity and specificity of different EPID systems using the same error detection criterion: a GAI threshold of 90 % at 3 %/2mm, as recommended by the AAPM TG 218 [29] as a universal action limit for IMRT QA. Additionally, we included two other criteria with larger (5 %) and smaller (2 %) dose differences to test sensitivity/specificity at different GAI levels. While this common criterion facilitates the comparison of results produced by different systems, it should not be interpreted as an endorsement or recommendation by our group for its universal use in error detection.

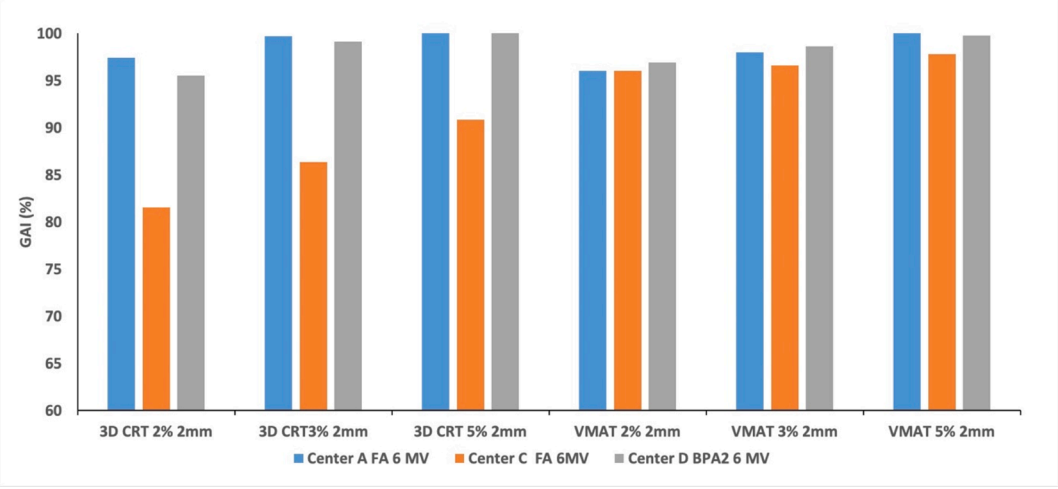

Fig. 1. Gamma agreement index (GAI) for the clinical plans (VMAT and 3DCRT) computed in an anthropomorphic lung phantom (CIRS thorax) by three centers.

Table 3
Sensitivity analysis for the 3DCRT plans. The first column shows the errors simulated in the TPS, with green indicating non-clinically significant errors and red indicating clinically significant errors. The columns to the right display the in vivo measurement results from different centers using various GAI criteria. The color codes are as follows: green for GAI $\geq$ 95 % and red for GAI $<$ 95 %.

Simulated Errors	Center/Algorithm/Energy	GAI criteria		
		2% 2mm	3% 2mm	5% 2mm
1cm bolus				
	Centre A FP 6 MV	86	96	100
	Centre C FP 6 MV	54	55	56
	Centre D BP2 6 MV	46	51	75
2 cm bolus				
	Centre A FP 6 MV	57	60	79
	Centre C FP 6 MV	53	53	53
	Centre D BP2 6 MV	33	33	34
3 cm bolus				
	Centre A FP 6 MV	46	47	48
	Centre C FP 6 MV	50	50	50
	Centre D BP2 6 MV	29	29	29
5 mm Lateral Shift				
	Centre A FP 6 MV	96	98	100
	Centre C FP 6 MV	63	66	70
	Centre D BP2 6 MV	97	99	100
10 mm Lateral Shift				
	Centre A FP 6 MV	90	94	100
	Centre C FP 6MV	70	74	82
	Centre D BP2 6 MV	84	90	97
15 mm Lateral Shift				
	Centre A FP 6 MV	84	88	96
	Centre C FP 6 MV	75	80	90
	Centre D BP2 6 MV	75	79	88
5 deg Coll rotation				
	Centre A FP 6 MV	98	100	100
	Centre C FP 6 MV	55	56	57
	Centre D BP2 6 MV	93	97	100
10 deg Coll rotation				
	Centre A FP 6 MV	100	100	100
	Centre C FP 6 MV	56	57	59
	Centre D BP2 6 MV	88	92	95
15 deg Coll rotation				
	Centre A FP 6 MV	99	100	100
	Centre C FP 6 MV	54	56	58
	Centre D BP2 6 MV	77	82	87

Error detection criteria should be optimized individually for each system, either through the use of receiver operating characteristic curves [28] or by applying process control theory to study normal process variability and establish appropriate confidence limits [29]. In this context, the error scheme proposed in our work may serve as a valuable tool for standardizing error definitions and optimize error detection. Although the errors tested in our study are not exhaustive, they encompass the most common failures that an in vivo system should detect, as detailed in the works of Olaciregui-Ruiz et al. [28] and Esposito et al. [19], and are easily implementable in most clinical settings.

For instance, while erroneous collimator rotation is relatively rare, it exemplifies a machine error (specifically, leaves misposition) that is frequently examined in the literature. Although modifying leaf positions in a DICOM file can be challenging and may not be easily implemented across all clinics, collimator rotation can be readily simulated in the TPS and implemented as a delivery error.“

A threshold of 30 % of the maximum dose for the GAI was chosen to align with various methodologies in the literature, facilitating better data comparison across tests. This threshold was a compromise based on different systems’ requirements. Typically, BP algorithms required higher threshold values: Esposito et al. used 30 % for Dosimetry Check [19], while Mijnheer et al. used 50 % [8]. Conversely, FP algorithms, as reported by Bossuyt et al. [20], employed a threshold of 10 %.

Table 4
Sensitivity analysis for the VMAT plans. The first column shows the errors simulated in the TPS, with green indicating non-clinically significant errors and red indicating clinically significant errors. The columns to the right display the in vivo measurement results from different centers using various GAI criteria. The color codes are as follows: green for GAI $\geq$ 95 % and red for GAI $<$ 95 %.

Simulated Errors	Center/Algorithm/energy	GAI criteria		
		2% 2mm	3% 2mm	5% 2mm
1 cm bolus				
	Centre A FP 6 MV	97	100	100
	Centre C FP 6 MV	86	86	87
	Centre D BP2 6 MV	69	78	90
2 cm bolus				
	Centre A FP 6 MV	88	88	89
	Centre C FP 6 MV	82	82	83
	Centre D BP2 6 MV	30	38	61
3 cm bolus				
	Centre A FP 6 MV	72	73	74
	Centre C FP 6 MV	80	81	81
	Centre D BP2 6 MV	14	17	32
5 mm L				
	Centre A FP 6 MV	97	98	100
	Centre C FP 6 MV	95	96	97
	Centre D BP2 6 MV	97	99	99
10 mm Lateral Shift				
	Centre A FP 6 MV	99	100	100
	Centre C FP 6 MV	98	99	100
	Centre D BP2 6 MV	95	98	99
15 mm Lateral Shift				
	Centre A FP 6 MV	100	100	100
	Centre C FP 6 MV	99	100	100
	Centre D BP2 6 MV	93	96	98
5 deg Coll rotation				
	Centre A FP 6 MV	94	97	100
	Centre C FP 6 MV	93	94	95
	Centre D BP2 6 MV	96	98	100
10 deg Coll rotation				
	Centre A FP 6 MV	90	93	95
	Centre C FP 6 MV	92	93	94
	Centre D BP2 6 MV	93	96	99
15 deg Coll rotation				
	Centre A FP 6 MV	85	87	89
	Centre C FP 6 MV	91	92	93
	Centre D BPA2 6 MV	85	90	96

Our results underscore several critical considerations. The performance of the same IVD system can vary significantly across different clinics, necessitating site-specific commissioning to address local factors. EPID performance may degrade over time due to portal imaging, requiring periodic recommissioning or recalibration to maintain accurate electronic responses [30,31]. The commissioning process for EPID algorithms is complex, often involving locally acquired data for beam configuration and dose calibration, which can be prone to errors [5–7,32,33]. Optimization of beam and EPID modelling procedures is crucial [34], as demonstrated by improved results in the BP2 algorithm over BP1, attributed to additional fields and measurements during configuration.

Energy dependence is another important factor, with different models required for each beam energy [5–7]. Sensitivity also varied by radiotherapy technique, with VMAT showing the lowest sensitivity compared to 3DCRT, likely due to compensatory effects in VMAT plans. A time-resolved comparison, rather than the time-integrated approach used, could potentially enhance sensitivity in VMAT plans [34].

Finally, all IVD systems struggle with small field dose distribution accuracy, particularly in SBRT and SRS, necessitating specific

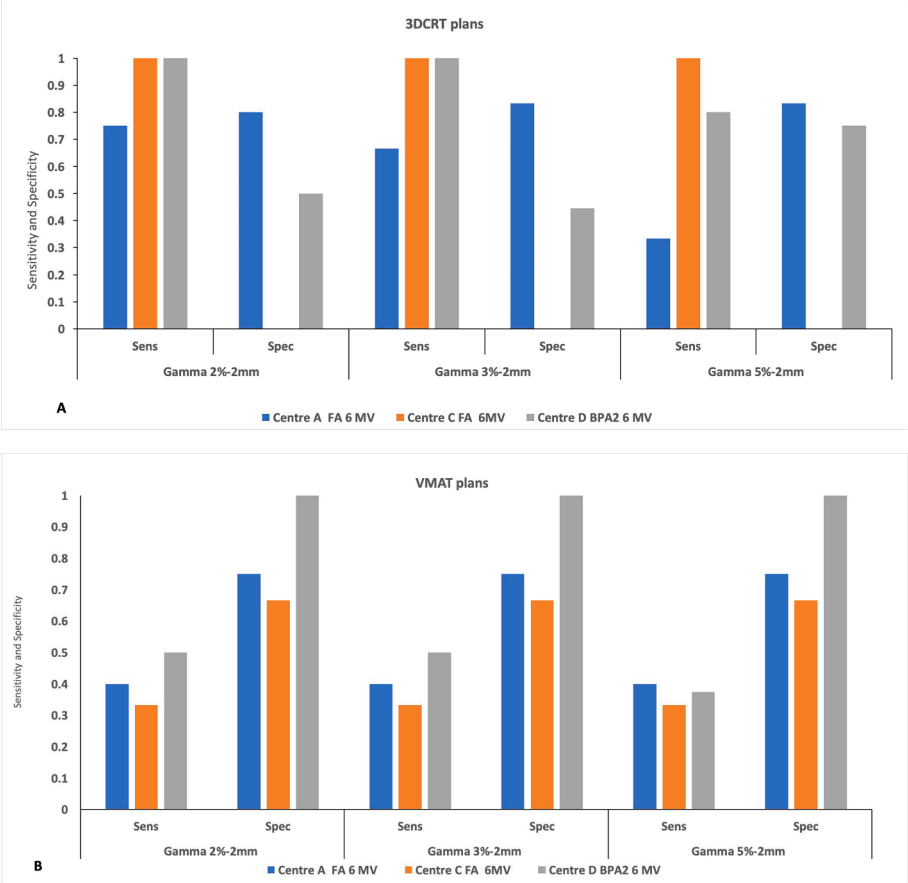

Fig. 2. Sensitivity and specificity for 3DCRT (a) and VMAT (b) plans for three different gamma criterion and for three Centers, A and C with a Forward Projection algorithm, Center D with a Back Projection algorithm.

calibration procedures [35,36], and potentially higher tolerance levels for dose differences while maintaining strict distance-to-agreement standards for safety.

In conclusion, we successfully developed a protocol for the commissioning of EPID IDV systems. During this process, we discovered that not all systems and their specific implementations reached the desired levels of accuracy and sensitivity. This highlights the critical importance of rigorous commissioning procedures and benchmarking to ensure optimal performance across different setups.

5. Author credit statement

Quality assurance of EPID transit dosimetry software: multicentric implementation of an independent evaluation protocol.“ Protocol design: Marco Esposito, Evy Bossuy, Catharine H Clark, Nuria Jornet, Stephen Kry, Jeroen van de Kamer, Dirk Verellen. Measurements: Marco Esposito, Riccardo Baldoni, Evy Bossuyt, Sara Brescia, Matthew Jones, Joseph Perry, Dirk Verellen Data Analysis: Marco Esposito Paper draft: Marco Esposito Nuria Jornet, Dirk Verellen Paper review and approval: all coauthors.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: The corresponding author is an Editorial Board Member for Physics and Imaging in Radiation Oncology and was not involved in the editorial review or the decision to publish this article. Núria Jornet is on the SunNuclear advisory board and SunNuclear reference site. Sara Brescani Candiolo Institute is a Sun Nuclear Reference site. Evy Bossuyt is on the SunNuclear advisory board and SunNuclear reference site. Dirk Verellen Iridium Netwerk, Antwerp is a SunNuclear reference site.

Acknowledgements

The authors would like to dedicate this manuscript to Ben Mijnheer who had a pivotal role in medical physics in Europe and a true pioneer in in vivo dosimetry and EPID-based IVD in particular

References

[1] World Health Organization (WHO) (2008) Radiotherapy risk profile.
[2] COUNCIL DIRECTIVE 2013/59/EURATOM of 5 December 2013. Available online: (accessed on https://eur-lex.europa.eu/eli/dir/2013/59/oj).
[3] Huyskens DP, Bogaerts R, Verstraete J, Lööf M, Nyström H, Jornet N, et al. Booklet 5: Practical Guidelines for the Implementation of in vivo dosimetry with diodes in external radiotherapy with photon beams (entrance dose). ESTRO publication 2001.
[4] International Atomic Energy Agency. Development of procedures for in vivo dosimetry in radiotherapy. IAEA, Vienna, Austria (IAEA Human Health Report No. 8); 2001.
[5] Dogan N, Mijnheer BJ, Padgett K, Nalichowski A, Wu C, Nyflot MJ, et al. AAPM Task Group Report 307: Use of EPIDs for Patient-Specific IMRT and VMAT QA. Med Phys 2023;50:e865–903. https://doi.org/10.1002/mp.16536.
[6] Esposito M, Villaggi E, Bresciani S, Cilla S, Falco MD, Garibaldi C, et al. Estimating dose delivery accuracy in stereotactic body radiation therapy: a review of in-vivo measurement methods. Radiother Oncol 2020;149:158–67. https://doi.org/10.1016/j.radonc.2020. 05.014.
[7] Olaciregui-Ruiz I, Beddar S, Greer P, Jornet N, McCurdy B, Paiva-Fonseca G, et al. In vivo dosimetry in external beam photon radiotherapy: requirements and future directions for research, development, and clinical practice. Phys Imaging Radiat Oncol 2020;15:108–16. https://doi.org/10.1016/j.phro.2020.08.003.
[8] Mijnheer BJ, González P, Olaciregui-Ruiz I, Rozendaal RA, van Herk M, Mans A. Overview of 3-year experience with large-scale electronic portal imaging device–based 3-dimensional transit dosimetry. Pract Radiat Oncol 2015;5: e679–87. https://doi.org/10.1016/j.prro.2015.07.001.

[9] Piermattei A, Greco F, Grusio M, Menna S, Azario L, Stimato G, et al. A validation study of a dedicated software for an automated in vivo dosimetry control in radiotherapy. Med Biol Eng Comput 2018;56:1939–47. https://doi.org/10.1007/s11517-018-1822-3.
[10] Celi S, Costa E, Wessels C, Mazal A, Fourquet A, Francois P. EPID based in vivo dosimetry system: clinical experience and results. J Appl Clin Med Phys 2016;17:262–76. https://doi.org/10.1120/jacmp.v17i3.6070.
[11] Nailon WH, Welsh D, McDonald K, Burns D, Forsyth J, Cooke G, et al. EPID-based in vivo dosimetry using Dosimetry Check™: Overview and clinical experience in a 5-yr study including breast, lung, prostate, and head and neck cancer patients. J Appl Clin Med Phys 2019;20:6–16. https://doi.org/10.1002/acm2.12441.
[12] Fiagan YAC, Bossuyt E, Machiels M, Nevens D, Billiet C, Poortmans P, et al. Comparing treatment uncertainty for ultra- vs. standard-hypofractionated breast radiation therapy based on in-vivo dosimetry. Phys Imaging Radiat Oncol 2022;22:85–90. https://doi.org/10.1016/j.phro.2022.05.003.
[13] Esposito M, Piermattei A, Bresciani S, Orlandini LC, Falco MD, Giancaterino S, et al. Improving dose delivery accuracy with EPID in vivo dosimetry: results from a multicenter study. Strahlenther Onkol 2021;197:633–43. https://doi.org/10.1007/s00066-021-01749-6.
[14] Bossuyt E, Nevens D, Weytjens R, Taieb Mokaddem A, Verellen D. Assessing the impact of adaptations to the clinical workflow in radiotherapy using transit in vivo dosimetry. Phys Imaging Radiat Oncol 2023;25:100420. https://doi.org/10.1016/j.phro.2023.100420.
[15] Esposito M, Mancosu P, Bruschi A, Ghirelli A, Pini S, Alpi P, et al. The role of EPID in vivo dosimetry in the risk management of stereotactic lung treatments. Strahlenther Onkol 2023;199:992–9. https://doi.org/10.1007/s00066-023-02081-x.
[16] Miri N, Lehmann J, Legge K, Vial P, Greer PB. Virtual EPID standard phantom audit (VESPA) for remote IMRT and VMAT credentialing. Phys Med Biol 2017;62:4293–9. https://doi.org/10.1088/1361-6560/aa63df.
[17] Kry SF, Molineu A, Kerns JR, Faught AM, Huang JY, Pulliam KB, et al. Institutional patient-specific IMRT QA does not predict unacceptable plan delivery. Int J Radiat Oncol Biol Phys 2014;90:1195–201. https://doi.org/10.1016/j.ijrobp.2014.08.334.
[18] Esposito M, Bruschi A, Bastiani P, Ghirelli A, Pini S, Russo S, et al. Characterization of EPID software for VMAT transit dosimetry. Australas Phys Eng Sci Med 2018;41:1021–7. https://doi.org/10.1007/s13246-018-0693-0.
[19] Esposito M, Marrazzo L, Vanzi E, Russo S, Pallotta S, Talamonti C. A Validation Method for EPID In Vivo Dosimetry Algorithms. Appl Sci 2021;11:10715. https://doi.org/10.3390/app112210715.
[20] Bossuyt E, Weytjens R, Nevens D, De Vos S, Verellen D. Evaluation of automated pre-treatment and transit in-vivo dosimetry in radiotherapy using empirically determined parameters. Phys Imaging Radiat Oncol 2020;16:113–29. https://doi.org/10.1016/j.phro.2020.09.011.
[21] INTERNATIONAL ATOMIC ENERGY AGENCY, Accuracy Requirements and Uncertainties in Radiotherapy, IAEA Human Health Series No. 31, IAEA, Vienna 2016.
[22] Prescribing, Recording, and Reporting Photon-Beam Intensity-Modulated Radiation Therapy (IMRT): Contents. Journal of the ICRU. 2010;10:1-3. https://doi.org/10.1093/jicru_ndq002.
[23] Bedford JL, Hanson IM, Hansen VN. Comparison of forward- and back-projection in vivo EPID dosimetry for VMAT treatment of the prostate. Phys Med Biol 2018;63:025008. https://doi.org/10.1088/1361-6560/aa9c60.
[24] Mijnheer B, Jomehzadeh A, González P, Olaciregui-Ruiz I, Rozendaal R, Shokrani P, et al. Error detection during VMAT delivery using EPID-based 3D transit dosimetry. Phys Med 2018;54:137–45. https://doi.org/10.1016/j.ejmp.2018.10.005.
[25] Yedekci Y, Biltekin F, Ozyigit G. Feasibility study of an electronic portal imaging based in vivo dose verification system for prostate stereotactic body radiotherapy. Phys Med 2019;64:204–9. https://doi.org/10.1016/j.ejmp.2019.07.008.
[26] Zhuang AH, Olch AJ. Sensitivity study of an automated system for daily patient QA using EPID exit dose images. J Appl Clin Med Phys 2018;19:114–24. https://doi.org/10.1002/acm2.12303.
[27] Bojechko C, Ford EC. Quantifying the performance of in vivo portal dosimetry in detecting four types of treatment parameter variations. Med Phys 2015;42:6912–8. https://doi.org/10.1118/1.4935093.
[28] Olaciregui-Ruiz I, Rozendaal R, Mijnheer B, Mans A. Site-specific alert criteria to detect patient-related errors with 3D EPID transit dosimetry. Med Phys 2019;46:45–55. https://doi.org/10.1002/mp.13265.
[29] 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-e83. Doi: 10.1002/mp.12810.
[30] Louwe RJ, McDermott LN, Sonke JJ, Tielenburg R, Wendling M, van Herk MB, et al. The long-term stability of amorphous silicon flat panel imaging devices for dosimetry purposes. Med Phys 2004;31:2989–95. https://doi.org/10.1118/1.1803751.
[31] McDermott LN, Nijsten SM, Sonke JJ, Partridge M, van Herk M, Mijnheer BJ. Comparison of ghosting effects for three commercial a-Si EPIDs. Med Phys 2006;33:2448–51. https://doi.org/10.1118/1.2207318.
[32] Hanson IM, Hansen VN, Olaciregui-Ruiz I, van Herk M. Clinical implementation and rapid commissioning of an EPID based in-vivo dosimetry system. Phys Med Biol 2014;59:171–9. https://doi.org/10.1088/0031-9155/59/19/N171.
[33] Olaciregui-Ruiz I, Osinga-Blaettermann JM, Ortega-Marin K, Mijnheer B, Mans A. Extending in aqua portal dosimetry with dose inhomogeneity conversion maps for accurate patient dose reconstruction in external beam radiotherapy. Phys Imaging Radiat Oncol 2022;22:20–7. https://doi.org/10.1016/j.phro.2022.04.001.
[34] Podesta M, Nijsten SM, Persoon LC, Scheib SG, Baltes C, Verhaegen F. Time dependent pre-treatment EPID dosimetry for standard and FFF VMAT. Phys Med Biol 2014;59:4749–68. https://doi.org/10.1088/0031-9155/59/16/4749.
[35] Parent L, Seco J, Evans PM, Fielding A, Dance DR. Monte Carlo modelling of a-Si EPID response: the effect of spectral variations with field size and position. Med Phys 2006;32:4527–40. https://doi.org/10.1118/1.2369465.
[36] Greer PB, Cadman P, Lee C, Bzdusek K. An energy-fluence convolution model for amorphous silicon EPID dose prediction. Med Phys 2009;36:547–55. https://doi.org/10.1118/1.3058481.