이학박사 학위논문

Studies on Tolerance Level and Sensitivity for Patient-specific Quality Assurance in Radiation Therapy

방사선치료 환자별 정도관리 허용오차 및 민감도에 관한 연구

2013 년 08 월

서울대학교 대학원

방사선응용생명과학 협동과정

김 정 인

A thesis of the Degree of Doctor of Philosophy

방사선치료 환자별 정도관리 허용오차 및 민감도에 관한 연구

Studies on Tolerance Level and Sensitivity for Patient-specific Quality Assurance in Radiation Therapy

August 2013

The Department of Interdisciplinary Program in

Radiation Applied Life Science,

Seoul National University

College of Medicine

Jung-in Kim

방사선치료 환자별 정도관리 허용오차 및 민감도에 관한 연구

지도교수 우 홍 균

이 논문을 이학박사 학위논문으로 제출함

2013 년 04 월

서울대학교 대학원

방사선응용생명과학 협동과정

김 정 인

김정인의 박사학위논문을 인준함

2013 년 07 월

위 원 장 김 희 찬 (인)

부위원장 우 홍 균 (인)

위 원 서 태 석 (인)

위 원 김 성 완 (인)

위 원 이 재 성 (인)

Studies on Tolerance Level and Sensitivity for Patient-specific Quality Assurance in Radiation Therapy

by
Jung-in Kim

A thesis submitted to the Department of Interdisciplinary Program in Radiation Applied Life Science in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Medical Physics at Seoul National University College of Medicine

July 2013

Approved by Thesis Committee:

Professor _____ **Chairman**
Professor _____ **Vice chairman**
Professor _____
Professor _____
Professor _____

학위논문 원문제공 서비스에 대한 동의서

본인의 학위논문에 대하여 서울대학교가 아래와 같이 학위논문 제공하는 것에 동의합니다.

1. 동의사항

① 본인의 논문을 보존이나 인터넷 등을 통한 온라인 서비스 목적으로 복제할 경우 저작물의 내용을 변경하지 않는 범위 내에서의 복제를 허용합니다.

② 본인의 논문을 디지털화하여 인터넷 등 정보통신망을 통한 논문의 일부 또는 전부의 복제.배포 및 전송 시 무료로 제공하는 것에 동의합니다.

2. 개인(저작자)의 의무

본 논문의 저작권을 타인에게 양도하거나 또는 출판을 허락하는 등 동의 내용을 변경하고자 할 때는 소속대학(원)에 공개의 유보 또는 해지를 즉시 통보하겠습니다.

3. 서울대학교의 의무

① 서울대학교는 본 논문을 외부에 제공할 경우 저작권 보호장치(DRM)를 사용하여야 합니다.

② 서울대학교는 본 논문에 대한 공개의 유보나 해지 신청 시 즉시 처리해야 합니다.

논문 제목: Studies on Tolerance Level and Sensitivity for Patient-specific Quality Assurance in Radiation Therapy

학위구분: **박사**
학 과: 방사선응용생명과학 협동과정
학 번: 2009-30597
연 락 처: madaning@gmail.com
저 작 자: 김 정 인 (인)

제 출 일: 2013 년 07 월 31 일

서울대학교총장 귀하

ABSTRACT

Purpose: The aim of this study was to investigate the tolerance level and sensitivity of patient-specific quality assurance (QA) for intensity modulated radiation therapy (IMRT) and volumetric modulated arc therapy (VMAT).

Methods: In order to investigate the tolerance level of patient-specific QA for IMRT in Korea, a multi-institutional study involving 12 radiation therapy institutions was performed. And then, based on the acquired tolerance level from multi-institutional study, sensitivities of various QA methods to detect errors in IMRT and VMAT plans were investigated. The multi-institutional study was performed by (1) point dose measurements using ion chamber at high- and low-dose regions (2) and planar (per-field and composite-field) dose measurements using film or 2-dimensional detector array. The multi-institutional study consisted of two programs, one was mock program and the other was clinic program. For mock program, we employed five mock structures reflecting anatomy of average Korean, while clinical treatment plans were used for clinic program to investigate the tolerance level. With the results of multi-institutional study, patient-specific QAs of point dose measurements with ion chamber, measurement of 2D dose distribution of axial plane with radiochromic film (EBT2), measurements of 2D dose distribution of coronal plane with MatriXX® and measurements of 3D volume dose distribution with COMPASS® has been performed. The results were compared with statistics by dividing into four groups according to delivery technique (IMRT group and VMAT group) and the degree of modulation (prostate group and H&N group).

Additionally, a new discretized deliverable VMAT plan (i.e., static arc (SA) plan) was generated and performed 3D QA to evaluate the dose discrepancy in planning target volume (PTV). One of the reasons possible to explain this discrepancy was an arc discretization of VMAT in treatment planning system (TPS). The dose discrepancy of PTV between VMAT and SA plans were evaluated using by gamma test with 3%/3 mm criteria and dose difference at 95% and 5% volume of PTV, respectively.

Results: Since the concept of confidence limit (CL) was appropriate for point dose measurement, tolerance level could be acquired. The tolerance level of point dose measurement at high-dose region was ±3% in both two programs. The tolerance level at low-dose region was ±5% in mock program, while it was ±7% and ±5% for linear accelerator (LINAC) and tomotherpy (TOMO) group in clinic program, respectively. On the contrary, for planar dose measurement, the concept of CL was not appropriate because of a large local deviation and a small number of samples. However, the results of planar dose measurement in both programs were well agreed with that of multi-institutional study performed by American Association of Physicists in Medicine (AAPM).

In the sensitivity study, the results of point dose measurement followed a normal distribution at all groups. The CL for IMRT, VMAT, prostate, and H&N groups were 3.0%, 2.1%, 1.0%, and 3.4%, respectively. The results of 2D dose measurements in axial and coronal plane showed significant differences at delivery groups (IMRT vs. VMAT) and patient groups (prostate vs. H&N) with the 3%/3 mm criteria. The results of 3D volume dose measurements showed significant differences at delivery groups and patient groups in both criteria.

There was no strong correlation between 2D QAs at axial plane and coronal plane. Similarly no correlation was observed between 2D and 3D QAs. Only 3D QA was possible to detect a dose discrepancy in PTV during delivery of VMAT. As the result of evaluation of arc discretization, the gamma passing rate of QA results were 92.1% and 96.8% for VMAT and SA QA, respectively. The dose differences at D_{95} were 2.61% and 0.97% and at D_5 were 2.71% and 0.04% for VMAT and SA QA, respectively.

Conclusions: Since the result of a multi-institutional study in Korea was coincident with those of AAPM and European Society for Therapeutic Radiology and Oncology (ESTRO) guidelines, patient-specific QA in Korea could be considered to meet the standards of international guideline. However, a point dose and 2D measurements recommended as patient-specific QA methods by international guidelines seem not to be enough to guarantee of an accurate delivery of IMRT and VMAT plans because no correlation was observed among point dose measurements, 2D QA and 3D QA. Three-dimensional QA was most sensitive to detect errors in treatment plans especially for VMAT. Therefore, it seems to be reasonable to adopt 3D QA as a patient-specific QA method for accurate radiation therapy using IMRT and VMAT techniques.

Keywords: Tolerance level, Sensitiviey, Quality assurance, Multi-institutional study, IMRT, VMAT

***Student number*: 2009-30597**

CONTENTS

LIST OF TABLES AND FIGURES

Chapter 1

Chapter 2

LIST OF ABBREVIATION

x%/x mm: x% dose difference and x mm distance to agreement

AAA: Analytical Anisotropic Algorithm

CL: Confidence Limit

COV: Coefficient of Variation

DVO: Dose Volume Optimizer

H&N: Head and Neck

IMRT: Intensity Modulated Radiation Therapy

LINAC: Linear Accelerator

MU: Monitor Units

OAR: Organ at Risk

OSLD: Optically Stimulated Luminescence Dosimeter

PTV: Planning Target Volume

QA: Quality Assurance

ROI: Region of Interest

SA: Static Arc

SAD: Source to Axis Distance

SD: Standard Deviation

SSD: Source to Surface Distance

STT: Segment Time Table

TOMO: Tomotherapy

TPS: Treatment Planning System

RMS: Root Mean Square

VMAT: Volumetric Modulated Arc Therapy

GENERAL INTRODUCTION

The primary goal of radiation therapy is to deliver a prescribed dose to a target volume while minimizing the deleterious effects of radiation to normal tissues [1-4]. IMRT is a radiation therapy technique in accord with this goal by modulating intensities of radiation *i.e*., by controlling multi-leaf collimators (MLC) [5-15]. Similarly, VMAT is a technique possible to modulate radiation intensities during a rotation of a gantry, which enables faster treatment with similar or better quality of IMRT [16-20]. Since IMRT and VMAT are complex techniques involving many steps and non-intuitive process, stringent patient-specific QA is needed for safe and accurate treatment. The tolerance level of patient-specific QA is difficult to quantify for general purposes, and often depends on a combination of prolonged institutional and individual experience and preferences. For that reason, AAPM TG-119 reported the tolerance levels of patient-specific QA by performing a multi-institutional study [21]. However, multi-institutional study on patient-specific QA reflecting the anatomy of average Korean has not been performed yet in Korea. Furthermore, even widely-accepted QA methods with tolerance levels recommended by international guidelines were not evaluated in depth in terms of sensitivity to detect errors. Even though few studies have been performed about the sensitivity of patient-specific QA, it is still unclear that a point dose measurement and a gamma evaluation of 2D QA could guarantee the accurate delivery of 3D dose distributions as planned [22-24]. In addition, even though the fundamental mechanism of VMAT in terms of operation and dose calculation is different from that of IMRT, same QA methods and tolerance

level of IMRT has been adopted for VMAT QA [25]. In this study, to investigate the tolerance level of patient-specific QA for IMRT in Korea reflecting anatomy of average Korean, a multi-institutional study involving 12 radiation therapy institutions was performed. And then, based on the acquired tolerance level from multi-institutional study, sensitivities of various QA methods with regard to error-detecting ability for IMRT and VMAT has been investigated by comprehensive comparison of patient-specific QA results and statistical analysis.

CHAPTER 1

A Multi-institutional Study on Tolerance Levels of IMRT Dose Quality Assurance Measurements in Korea

INTRODUCTION

IMRT is a radiation therapy technique possible to provide a conformal prescribed dose to a complex-shaped target while sparing normal tissue by intensity modulation [12]. Since IMRT is a complex technique involving many steps and non-intuitive process, a stringent patient-specific QA is needed for a safe and accurate delivery of a treatment [26-31]. The AAPM and ESTRO emphasized a comprehensive QA program for clinical implementation of IMRT [32, 33]. Also, the guideline of IMRT for clinical trials has been developed throughout the world [34, 35].

The tolerance level of a patient-specific QA is difficult to quantify for general purposes, and often depends on a combination of prolonged institutional and individual experience and preferences. AAPM Task Group 119 carried out a multi-institutional study to assess the overall accuracy of planning and delivery of IMRT, and produced quantitative confidence limits as baseline expectation values for IMRT commissioning [36]. The British group carried out a national dosimetry audit of IMRT to provide an independent check of safe implementation and to identify problems in the modeling and delivery of IMRT [35, 37-39].

In spite of the rapid increase of IMRT in Korea, no multi-institutional study about patient-specific IMRT QA in terms of tolerance level has been performed yet. Radiation therapy institutions in Korea establish their own IMRT QA protocols based on international guidelines such as AAPM TG reports. Therefore, there is no national safety guideline in Korea and the domestic quality of patient-specific QA is hard to be verified. Moreover, there are

anatomical differences between westerner and Korean potentially possible to affect the IMRT plan even though the differences are not large. For that reason, in order to investigate the tolerance level of patient-specific QA for IMRT in Korea reflecting anatomy of average Korean, a multi-institutional study involving 12 radiation therapy institutions was performed. The tolerance levels of patient-specific IMRT QA as a national reference data for overall accuracy of IMRT planning and delivery was provided. The confidence limit concept and test protocol of AAPM TG-119 were adopted with modification and the data from 12 radiation therapy institutions has been analyzed statistically. This study was the first multi-institutional study on patient-specific QA in Korea.

MATERIALS AND METHODS

This study was performed from October 2010 to September 2012. Twelve participating institutions in Korea were divided into the linear accelerator group (LINAC) and tomotherapy group (TOMO) according to the treatment equipment. The planning and delivery systems of each institution are summarized in Table 1.1. Both groups used 6 MV photon beams. The institutions listed in subsequent tables were anonymously identified only by letter.

Table 1.1. Lists of participating institutions and the planning and delivery systems used

Group	Institution	Accelerator	Delivery technique	Planning system
LINAC	Seoul Nat'l Univ. Bundang Hosp.	Varian 21ExS	DMLC	Eclipse 6.5
	Jeju Nat'l Univ. Hosp.	Varian IX	DMLC	Eclipse 8.6
	Yeungnam Univ. Hosp.	Varian 21ExS	DMLC	Eclipse 8.6
	Dong-A Univ. Hosp. (only mock program)	Varian Novalis	DMLC	BrainLab iPlan
	Eulji Univ. Hosp.	Elekta Synergy	SMLC	CMS Monaco 2.0.3
	Seoul Nat'l Univ. Hosp.	Varian IX	DMLC	Eclipse 8.9
	Asan Medical Center	Varian Trilogy	DMLC	Eclipse 8.9
	Kangbuk Samsung Medical center (only clinic program)	Varian IX	DMLC	Eclipse 8.9
	VMS (only clinic program)	Varian IX	DMLC	Eclipse 8.9
TOMO	Seoul Samsung Medical Center	Tomotherapy	BMLC	Tomotherapy TPS 3.1.4
	Yonsei Cancer Center	Tomotherapy	BMLC	Tomotherapy TPS 4.0.2
	Chonnam Nat'l Univ. Hwasun Hosp.	Tomotherapy	BMLC	Tomotherapy TPS 3.2.3.2

Abbreviations: DMLC, SMLC, and BMLC stand for dynamic MLC, static MLC, and binary MLC.

1. Mock Program

The AAPM TG-119 used mock structures for the prostate, H&N, C-shape, and multi-target test. Each test included targets, normal structures, planning specifications of dose goals and beam arrangements. This study adopted the general guideline of the AAPM TG-119 programs. However, the target volumes and organ at risk (OAR) locations of the mock prostate and H&N were modified based on Korean patients' anatomy [40]. The rest of the AAPM TG-119 mock structures were applied identically for this study.

All mock structures were segmented by one physicist from the reference site on DICOM CT images. Then, these DICOM files were distributed to the institutions participating in this study to eliminate any institutional variations during segmentation. It was recommended that a grid size of dose calculation should be less than 3 mm and that the calculation algorithm should be used convolution–superposition or equivalent algorithm for inhomogeneity correction [41, 42].

For prostate test, the mean PTV and CTV volumes of sample Korean patients were 141 cc and 50 cc, respectively. The PTV was defined to include a 1.0 cm margin around the prostate in all directions, except the posterior direction where a 0.5 cm margin was added. The rectum was a cylinder with a diameter of 1.5 cm (mean volume of 10 cc) and the bladder was a semi-ellipsoidal shape (mean volume of 144 cc). The PTV included about one-third of the rectum and bladder volumes. Unlike AAPM TG-119, the femoral heads of spherical shape were added in this test suite.

For H&N test, the mean PTV volume of sample Korean patients was 534 cc. PTV included all anterior volume from the base of the skull to the upper neck and the posterior neck node. Both parotids were delineated with two "*truncated cones" (i.e.*, a cone with the top cut off) of 1.5 cm diameter of a circular top and 2.4 cm diameter of a circular bottom opposing the circular bottom with 30 cc volume. They were located at the superior aspect of the PTV. The cord was a cylinder shape with a diameter of 1.5 cm. A gap between the cord and PTV was about 1.3 cm.

For the C-shape and multi-target tests, the AAPM TG-119 structures were used as is. The mock structures are shown in Figure 1.1.

For the prostate and multi-target tests, the plan had seven fields at 50° intervals from the vertical (0°, 50°, 100°, 150°, 210°, 260°, and 310°). For H&N and C-shape tests, the plan had nine fields at 40° intervals from the vertical (0°, 40°, 80°, 120°, 160°, 200°, 240°, 280°, and 320°). The total prescribed dose was 80 Gy (2 Gy per fraction) for the prostate test while the total prescribed dose of 50 Gy (2 Gy per fraction) was applied for H&N, C-shape, and multi-target tests. The plan goals of C-shape were divided into easy and hard versions. The specific planning goals are shown in Table 1.2.

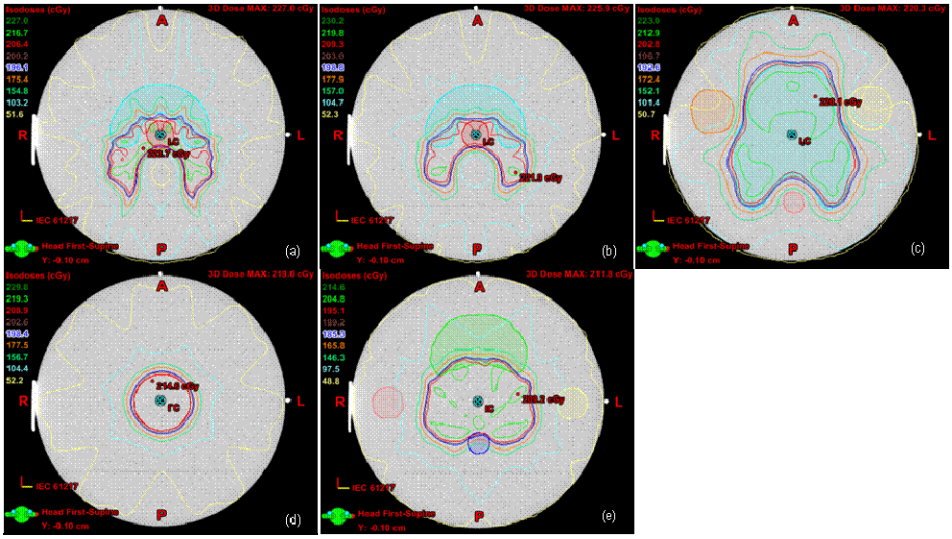

Figure 1.1. Isodose line for high dose point measurements and composite field plane with mock structures; (a) C-shape easy, (b) hard, (c) H&N, (d) Multi-target center, and (e) Prostate.

Table 1.2. Treatment plan goals and results for all tests of both groups in mock program

Group			LINAC			TOMO		
Test	Parameter	Goal	Mean	SD	COV	Mean	SD	COV
Prostate	PTV D_{95}	>7600	7620.4	52.9	0.007	7788.3	166.3	0.021
	PTV D_{5}	<8400	8267.4	188.1	0.023	8111.0	118.3	0.015
	Rtm D_{30}	<7000	6630.6	392.4	0.059	6270.0	628.6	0.100
	Rtm D_{10}	<7500	7324.6	208.7	0.028	7694.0	146.9	0.019
	Bld D_{30}	<7000	5452.7	738.7	0.135	5346.7	660.3	0.124
	Bld D_{10}	<7500	7414.8	144.8	0.020	7729.3	233.2	0.030
	RtF D_{10}	<5000	4127.4	503.4	0.122	4171.7	926.4	0.222
	LtF D_{10}	<5000	4014.8	504.5	0.126	4112.7	768.7	0.187
H&N	PTV D_{90}	5000	5052.9	84.6	0.017	4996.7	70.2	0.014
	PTV D_{99}	>4650	4784.1	94.8	0.020	4883.3	98.7	0.020
	PTV D_{20}	<5500	5289.6	147.2	0.028	5215.7	149.3	0.029
	Cd max	<4000	3915.0	257.5	0.066	3282.0	499.2	0.152
	RtPd D_{50}	<2000	1916.7	160.2	0.084	1438.3	192.4	0.134
	LtPd D_{50}	<2000	1887.3	135.3	0.072	1394.3	144.7	0.104
C-shape(E)	PTV D_{95}	5000	4985.6	64.8	0.013	4982.0	19.3	0.004
	PTV D_{10}	<5500	5463.3	188.6	0.035	5437.7	207.7	0.038
	Core D_{10}	<2500	2446.3	145.1	0.059	1793.3	583.2	0.325
C-shape(H)	PTV D_{95}	5000	4937.0	116.6	0.024	4790.7	116.8	0.024
	PTV D_{10}	<5500	5639.4	162.7	0.029	5950.0	578.2	0.097
	Core D_{10}	<1000	1552.9	211.9	0.136	1178.3	361.4	0.307
Multiple-target	Ct D_{99}	>5000	4975.6	54.0	0.011	4918.0	74.6	0.015
	Ct D_{13}	<5300	5417.2	117.1	0.022	5852.7	1002.6	0.171
	Sup D_{99}	>2500	2676.2	204.2	0.076	2437.3	84.5	0.035
	Sup D_{13}	<3500	3521.3	352.0	0.100	3700.7	149.9	0.041
	Inf D_{99}	>1250	1430.3	353.2	0.247	1220.0	60.8	0.050

Inf D_{13}	<2500	2593.5	607.2	0.234	3262.7	423.8	0.130

Abbreviations: PTV, Rtm, Bld, RtF, LtF, Cd, Ct, Sup, Inf, SD, COV, E, and H stand for planning target volume, rectum, bladder, right femoral head, left femoral head, cord, center, superior, inferior, standard deviation, coefficient of variation, easy, and hard.

2. Clinic Program

Eleven institutions were grouped into 8 LINAC and 3 TOMO. Clinical IMRT cases from each institution included brain, abdomen, H&N, and prostate. The total number of cases for brain, H&N, abdomen, and prostate were 20, 60, 18, and 57, respectively. The structure, prescription, goals, and parameters for planning were entirely dependent on the strategies of local institution. However, it was recommended that the grid size and dose calculation algorithm should be used same as mock program.

3. Custom-made Cylindrical Phantom

Patient-specific QA in LINAC group performed with a same phantom. The phantom was custom-made and made of acrylic. The cylindrical phantom was 265 mm in length and 180 mm in diameter. It had two holes for the insertion of an ion chamber. It was cut into two pieces for insertion of film as shown in Figure 1.2. The two pieces of phantom (gray and green parts in Figure 1.2) were tightened using a lever after inserting a film in order to reduce the air gap. One hole at 5 cm depth below the anterior surface was used to measure a conversion factor [nC/cGy] of chamber reading-to-dose. This standard measurement was also intended to exclude the daily variation of machine output. The other hole along the axis was used to measure a planned dose at a high- or low-dose region. This phantom was designed for both point dose measurement and 2D dose distribution measurement with film.

For TOMO group, the commercial phantom (cheese phantom, Accuray, Sunnyvale, CA) was used to measure point doses and coronal plane dose

distributions. Details of the measurements with this phantom are described elsewhere [43].

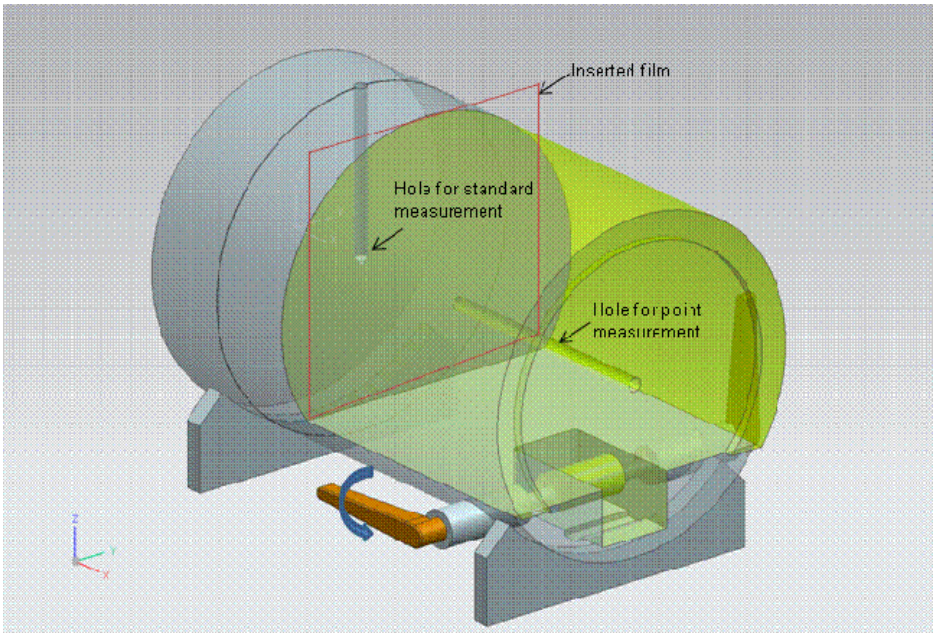

Figure 1.2. Developed IMRT DQA phantom for LINAC group.

4. Output and TPS audit

One of the institutions has been participating in a radiation therapy oncology group (RTOG) trial and passed the RPC's (Radiological Physics Center, MD Anderson Cancer Center, USA) credentialing requirements. The results of RPC auditing guaranteed that the output deviation of this institution was less than ±2%. This institution played a role as a reference site to indirectly verify the machine output of the other institutions. The output of each institution was measured by skilled physicists of the reference site using a farmer type ionization chamber (0.6 cc, PTW TN30013; PTW, Freiburg, Germany) and an electrometer (PTW UNIDOS) calibrated by Korea Research Institute of Standards and Science (KRISS).

The treatment planning system commissioning audit was carried out using two test protocols described by Van Esch *et al.* [35, 44]. These tests were designed with three consecutive rectangular volumes that had different specified doses. The first test called a "*Dip*" test, for which the specified dose to the middle volume and each outer volume was 0.7 and 0.0 Gy, respectively. This test was performed for dynamic delivery to verify that the leaves could adequately shield the central volume, and the TPS modeled the transmission of the leaves correctly. The second test is called a "*Step*" test, for which the specified dose to each volume was 0.7, 0.5, and 0.3 Gy, respectively. The Step test was for static delivery to test the delivered accuracy of three relative dose levels. Each institution delineated the predefined volumes for both tests on a local solid water phantom and delivered the specified doses to films (EBT2, International

Specialty Products, Wayne, NJ). The reference site centrally evaluated the films from all institutions.

5. Dose Quality Assurance Measurements for Mock Program

LIANC group used a custom-made phantom with 0.125 cc ion chamber (Semiflex, PTW, Freiburg, Germany) and TOMO group used cheese phantom with a 0.05 cc ion chamber (Exradin A1SL, Standard Imaging Inc., Middleton, WI) to measure a point dose. A location of high-dose measurement was at the isocenter in the middle of PTV, where doses were high and uniform (see Figure 1.1). A location of low-dose measurements was in the OAR structure. The location of low-dose measurements were 3 cm posterior to the rectum for the prostate, 4 cm posterior in the middle of spinal cord for H&N, the center of cord for the C-shape and the center of either of two outer targets for the multi-target (sees Figure 1.3).

For per-field measurements, a plane perpendicular to the beam axis located at 5 cm depth in a water-equivalent phantom with the source-to-axis distance (SAD) setup was chosen to be compared with calculated dose distribution. Each institutions locally used an available 2D detector. The per-field measurement was only performed for H&N test and limited to the LINAC group. The individual fields were delivered at gantry angle of 0° to the angular dependency of detectors during per-field measurements [45, 46].

For composite field measurements, an iso-plane perpendicular to the beam axis was chosen for 2D QA. All institutions were asked to perform the composite field measurements using films and the custom-made (LINAC) or cheese (TOMO) phantom. The multi-target test was excluded from this

measurement. The films measurements were evaluated using two gamma criteria of 2%/2 mm and 3%/3 mm. The planar dose distributions were normalized at a reference (dose to the isocenter) or maximum dose in a low-gradient region. The region of interest (ROI) was first specified as a maximum size of rectangle on the film. Then any pixels that received less than 10% of the maximum dose in the dose map were excluded from the gamma evaluation.

6. Dose Quality Assurance Measurements for Clinic Program

For point dose measurement, a location of high-dose measurement was at the isocenter in the middle of PTV while a location of 30% - 50% of prescribed dose was selected as a point of low-dose measurements. The measurement instrument was identical to mock program.

The per-field measurements at the gantry angle of 0º were performed according to each institution's QA protocol and limited to the LINAC group.

The composite-field dose measurements were also performed following the local institution strategies. Each institution in the LINAC group used a detector array; two institutions used MapCHECK®; one institution used the ion chamber array; four institutions used MatriXX®; and one institution used ArcCheck® (Sun Nuclear Corporation, Melbourne, FL). TOMO group used a film.

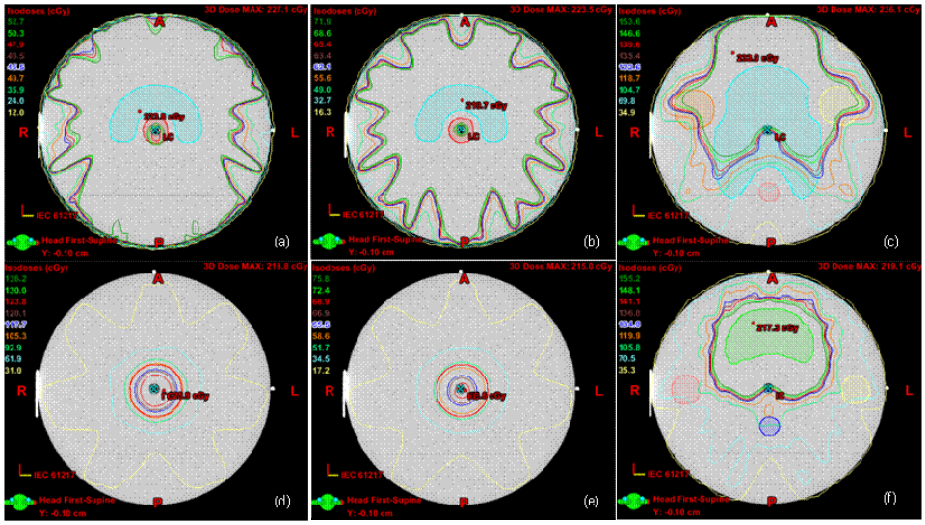

Figure 1.3. Isodose lines for low-dose point measurement with mock structures: (a) C-shape easy, (b) C-shape hard, (c) H&N, (d) superior of multi-target, (e) inferior of multi-target, and (f) prostate.

RESULTS

Planning in the Mock Program

The statistics of the mock plans for both groups were presented in Table 1.2. In this table, D_n means the dose covering n% of the volume. The coefficient of variation (COV) was a normalized measurement of the dispersion of a probability distribution that was defined as a ratio of the SD to the mean.

Output and TPS Audit

The results of output audit for all participating institutions ranged from -1.8% to +2.4%, which indicated that all of them passed the criteria of less than 3%. Six institutions undertook "*Dip and Step*" tests for the TPS audit, and all test plans met the required dose constraints. Since the other institutions did not perform these tests, they were requested to present the results of local TPS commission tests with film measurements. The average percentage of points passing the gamma criteria of 3%/3 mm in the "*Dip*" and "*Step*" tests were 98.2% (from 97.4% to 99.2%) and 97.8% (from 97.2% to 98.6%), respectively.

Point dose measurements

In this study, the dose difference was expressed as a ratio of the difference between measured and calculated doses to the calculated dose, instead of the prescription dose used in AAPM TG-119. The dose difference (%) was defined as

$$\text{Dose difference}\ (\%) = \frac{Measured\ dose - Planned\ dose}{Planned\ dose} \times 100\ (\%)$$

The calculated dose from TPS was a value at a point of measurement for comparison. In mock program, the average difference between measured and planned doses averaged over all tests for high dose measurement was -0.7% ± 1.2% at LINAC, -0.5% ± 1.4% at TOMO and -0.6% ± 1.3 at all institutions. It ranged from -3.3% to 1.9% at the LINAC group and from -2.5% to 2.9% at the TOMO group. The maximum dose difference occurred in the hard C-shape structure at both groups. The average confidence limit was 3.1% at both groups. The results of these measurements for both groups are summarized in Table 1.3. In low-dose measurements, the average difference between measured and calculated doses was -1.0% ± 1.9% at LINAC group, 0.1% ± 2.5% at TOMO group, and -0.6% ± 2.2% at all institutions. It ranged from -6.0% to 3.5% at the LINAC group and from -3.4% to 6.6% at TOMO group. The maximum dose difference also occurred in the hard C-shape structure at both groups. The average confidence limit was 4.9% at both groups. The results of these measurements are shown in Table 1.4. The local confidence limit of each institution, averaged over all the test plans, is listed in Table 1.5.

In clinic program, for high dose point measurements, the average difference between measured and calculated doses was -0.1% ± 1.2% at LINAC, -0.1% ± 1.4% at TOMO and -0.1% ± 1.3% at all institutions. It ranged from -3.1% to 3.0% at the LINAC group and from -2.8% to 2.9% at the TOMO group. These results are shown in Table 1.6. In low-dose measurements, the average difference between measured and calculated doses was –1.3% ± 2.6% at LINAC group, -0.9% ± 2.5% at TOMO group and -1.2% ± 2.5% at all institutions. It ranged from -9.4% to 5.0% at the LINAC group and from -5.5%

to 8.9% at TOMO group. These results are shown in Table 1.7. The local confidence limit of each institution is listed in Table 1.8.

Table 1.3. Results of high-dose point measurement, averaged over the institutions for LINAC/TOMO/Total groups in mock program

Test	Mean (%)	SD (%)	CL (%)	N
Multi-target	-0.3/-0.2/-0.2	1.0/1.8/1.2	2.3/3.6/2.5	7/3/10
Prostate	0.0/-0.6/-0.3	1.0/1.2/1.1	1.9/3.0/2.4	7/6/13
H&N	-1.0/-1.3/-1.1	0.7/0.8/0.7	2.4/2.8/2.6	7/6/13
C-shape (E)	-1.3/0.2/-0.9	1.1/1.4/1.4	3.5/3.0/3.5	7/3/10
C-shape (H)	-0.8/0.5/-0.4	1.9/2.2/2.0	4.5/4.7/4.2	7/3/10
Overall combined	-0.7/-0.5/-0.6	1.2/1.4/1.3	3.1/3.2/3.1	35/21/56

Abbreviations: SD, CL, E, and H stand for standard deviation, confidence limit, easy, and hard.

Table 1.4. Results of low-dose point measurement, averaged over the institutions for LINAC/TOMO/Total groups in mock program

Test	Mean (%)	SD (%)	CL (%)	N
MT (sup)	-0.3/1.0/0.1	1.8/2.7/2.1	3.9/6.3/4.1	7/3/10
MT (inf)	-0.5/-0.1/-0.4	1.6/3.0/1.9	3.3/6.0/4.0	7/3/10
Prostate	-1.5/-1.0/-1.2	1.3/1.2/1.2	3.9/3.4/3.6	7/6/13
H&N	-1.7/-0.5/-1.1	0.8/2.1/1.6	3.3/4.5/4.2	7/6/13
C-shape (E)	0.2/-0.7/-0.1	2.3/2.6/2.3	4.8/5.7/4.6	7/3/10
C-shape (H)	-2.3/3.7/-0.5	2.6/2.6/3.8	7.4/8.7/7.9	7/3/10
Overall	-1.0/0.1/-0.6	1.9/2.5/2.2	4.8/5.0/4.9	42/24/66

Abbreviations: SD, CL, E, and H stand for standard deviation, confidence limit, easy, and hard.

Table 1.5. Results of dose point measurement, averaged over all the test plans measured at each in mock program

Measurement	Group	LINAC						TOMO			
	Institution	A	B	C	D	E	F	G	H	I	J
High Dose	Mean (%)	-1.0	-1.1	-0.4	0.3	-1.8	-0.3	-0.5	-1.1	-0.1	-0.1
	SD (%)	0.8	1.6	1.2	0.8	1.4	0.5	1.5	1.3	1.7	1.1
	LCL (%)	2.6	4.3	2.7	1.9	4.5	1.2	3.4	3.6	3.4	2.3
	N	5	5	5	5	5	5	5	7	7	7
Low Dose	Mean (%)	-1.5	-1.7	0.4	-0.7	-2.2	-1.7	0.4	-0.6	1.1	-0.1
	SD (%)	1.0	2.4	1.7	1.8	2.0	1.3	1.8	2.1	3.3	1.6
	LCL (%)	3.4	6.4	3.7	4.1	6.2	4.3	4.0	4.7	7.7	3.3
	N	6	6	6	6	6	6	6	8	8	8

Abbreviations: SD and LCD stand for standard deviation and local confidence limit.

Table 1.6. Results of high-dose point measurement, averaged over the institutions for LINAC/TOMO/Total groups in clinic program

Test	Mean (%)	SD (%)	CL (%)	N
Brain	0.2/-0.7/-0.1	1.4/1.1/1.3	2.9/2.7/2.7	14/6/20
H&N	-0.2/0.2/0.0	1.7/1.4/1.5	3.5/2.9/3.0	22/38/61
Abdomen	0.0/-0.9/-0.4	1.1/1.6/1.4	2.2/3.9/3.1	10/8/18
Prostate	-0.2/0.0/-0.2	1.0/1.5/1.1	2.1/3.0/2.3	45/12/57
Overall combined	-0.1/-0.1/-0.1	1.2/1.4/1.3	2.5/2.9/2.7	91/65/156

Abbreviations: SD and CL stand for standard deviation and confidence limit.

Table 1.7. Results of low dose point measurement with the range of 30~50%, averaged over the institutions for LINAC/TOMO/Total groups in clinic program

Test	Mean (%)	SD (%)	CL (%)	N
Brain	-1.9/-2.6/-2.1	4.5/0.9/3.7	10.6/4.3/9.3	11/5/16
H&N	-1.2/-0.8/-0.9	2.9/2.1/2.4	6.9/4.9/5.6	12/24/36
Abdomen	-1.7/0.1/-0.9	2.0/3.3/2.7	5.7//6.5/6.2	8/7/15
Prostate	-1.3/-1.2/-1.3	2.1/1.7/2.0	5.4/4.6/5.3	36/6/42
Overall combined	-1.4/-0.9/-1.2	2.7/2.2/2.5	6.7/5.3/6.2	67/42/109

Abbreviations: SD and CL stand for standard deviation and confidence limit.

Table 1.8. Results of dose point measurement, averaged over all the test plans measured at each in clinic program

Measurement	Group	LINAC								TOMO		
	Institution	A	B	C	D	E	F	G	H	I	J	K
High Dose	Mean (%)	-0.5	2.2	-0.8	0.0	-0.2	-0.2	-0.3	-0.4	1.0	-0.8	-0.3
	SD (%)	0.3	0.8	0.4	1.7	0.6	1.8	0.9	1.1	1.0	1.4	1.3
	LCL (%)	1.2	3.8	1.7	3.3	1.4	3.6	2.1	2.4	2.9	3.6	2.8
	N	9	7	10	12	20	10	11	12	20	20	24
Low Dose	Mean (%)	-1.8	0.9	-1.6	-3.2	-2.4	-0.7	0.1	-1.3	-0.4	0.6	-2.1
	SD (%)	2.8	4.0	0.2	5.1	2.8	1.8	1.4	1.3	2.1	3.0	1.2
	LCL (%)	7.3	8.7	2.0	1.3	7.8	4.2	2.8	3.9	4.6	6.5	4.3
	N	9	5	3	4	20	9	8	9	14	9	19

Abbreviations: SD and LCL stand for standard deviation and local confidence limit.

Planar Dose Measurements in mock program

For per-field measurements, the average passing rate with the gamma criteria of 2%/2 mm and 3%/3 mm was 92.7% ± 6.5% and 98.2% ± 2.8%, respectively. The corresponding confidence limit was 79.1% and 92.7%. The local confidence limit ranged from 71.4% to 95.3% with 2%/2 mm criteria and from 88.3% to 100% with 3%/3 mm criteria as shown in the Table 1.9.

For composite field measurements, Table 1.10 summarized the passing rate averaged over all institutions of both groups and the associated confidence limits. The gamma passing rate averaged over all mock test plans in LINAC group was 84.7% ± 7.5% with 2%/2 mm criteria and 94.6% ± 4.0% with 3%/3 mm criteria while it was 88.4% ± 3.7% with 2%/2 mm criteria and 96.4% ± 3.2% with 3%/3 mm criteria in TOMO group. The gamma passing rate averaged over all mock test plans in all institutions was 86.1% ± 6.5% with 2%/2 mm criteria and 95.3% ± 3.8% with 3%/3 mm criteria. TOMO group showed a higher passing rate and a lower standard deviation than LINAC group. The local confidence limit is summarized in Table 1.11.

Planar Dose Measurements in clinic program

For per-field measurements, the average gamma passing rate of all tests was 92.4 ± 7.4% and 97.5 ± 3.4% with 2%/2 mm and 3%/3 mm criteria, respectively. The corresponding confidence limit was 77.9% and 90.8% as shown in the Table 1.12. The local confidence limit ranged from 88.7% to 97.1% with 2%/2 mm criteria and from 96.1% to 99.1% with 3%/3 mm criteria as shown in the Table 1.13.

For composite field measurements, the average gamma passing rate for LINAC group was 92.9 ± 7.4% and 98.3 ± 1.9% with 2%/2 mm and 3%/3 mm criteria, respectively. The average gamma passing rate for TOMO group was 89.6 ± 7.4% and 97.2 ± 2.3% with 2％/2 mm and 3%/3 mm criteria, respectively. These results are shown in the Table 1.14. The local confidence limit is summarized in Table 1.15.

Table 1.9. Per-field measurement: local and total averaged percentage of points passing gamma criteria of 2%/2 mm and 3%/3 mm, with associated confidence limits in mock program

	Institution	A	B	C	D	E	F	G	
Criteria	Device	Map-CHECK	Map-CHECK	EBT2	MatriXX	EBT2	2D-array	MatriXX	Total
2%/2 mm	Mean (%)	95.2	93.0	90.5	96.3	79.1	96.8	98.1	92.7
	SD (%)	2.9	1.7	5.4	1.0	4.0	2.7	1.5	6.5
	LCL (%)	89.4	89.7	80.0	94.4	71.4	91.5	95.3	79.1
3%/3 mm	Mean (%)	99.4	99.3	98.0	99.0	92.4	99.0	100.0	98.2
	SD (%)	1.0	0.5	2.1	0.9	2.1	1.3	0.0	2.8
	LCL (%)	97.4	98.3	93.9	97.2	88.3	96.5	100	92.7
	N	9	9	9	9	9	9	9	63

Abbreviations: SD and LCL stand for standard deviation and local confidence limit.

Table 1.10. Composite film: percentage of points passing gamma criteria of 2%/2 mm and 3%/3 mm, with associated confidence limits for LINAC/TOMO/Total groups in mock program

Criteria	Test	Prostate	H&N	C-shape (E)	C-shape (H)	Overall
2%/ 2 mm	Mean (%)	86.5/88.1/87.2	85.2/86.6/85.7	85.4/92.1/88.8	81.7/88.6/83.2	84.7/88.4/86.1
	SD (%)	6.1/2.9/4.8	7.3/0.9/5.7	8.3/7.4/8.2	9.1/3.2/8.5	7.5/3.7/6.5
	CL (%)	74.6/82.4/77.9	70.9/84.8/74.5	69.1/77.6/72.7	63.9/82.4/66.6	69.9/81.1/73.3
3%/ 3 mm	Mean (%)	95.4/96.4/95.8	94.6/96.1/95.3	95.1/98.0/96.0	93.3/95.5/94.0	94.6/96.4/95.3
	SD (%)	3.4/3.6/3.4	3.4/4.2/3.7	4.1/1.5/3.7	5.3/2.1/4.6	4.0/3.2/3.8
	CL (%)	88.7/89.3/89.2	87.9/87.8/88.0	87.1/95.1/88.7	82.9/91.3/85.0	86.8/90.0/87.9
	N	7/6/13	7/6/13	7/3/10	7/3/10	28/18/46

Abbreviations: SD and CL stand for standard deviation and confidence limit.

Table 1.11. Composite film: percentage of points passing gamma criteria of 2%/2 mm and 3%/3 mm, averaged over the test plans, with associated confidence limits in the mock program

Criteria	Group	LINAC							TOMO		
	Institution	A	B	C	D	E	F	G	H	I	J
2%/2 mm	Mean (%)	91.1	81.0	77.2	84.1	78.9	95.4	85.0	86.3	89.0	89.8
	SD (%)	4.3	1.3	4.8	6.9	8.3	2.5	0.9	2.2	4.2	4.1
	LCL (%)	82.6	78.5	67.7	70.6	62.6	90.4	83.2	82.0	80.8	81.8
3%/3 mm	Mean (%)	98.0	93.1	90.1	95.4	90.7	99.3	95.7	96.6	93.8	98.8
	SD (%)	1.3	0.7	3.2	3.4	4.3	0.6	1.2	3.0	2.9	1.7
	LCL (%)	95.5	91.7	83.9	88.7	82.2	98.1	93.4	90.7	88.1	95.5
	N	4	4	4	4	4	4	4	6	6	6

Abbreviations: SD and LCL stand for standard deviation and local confidence limit.

Correlation between the magnitude of point dose error and gamma passing rates of composite field measurements in mock program

The absolute point dose error (%) is defined as

$$\text{Absolute point dose error (\%)} = \left|\frac{Measured\ dose - Planned\ dose}{Planned\ dose}\right| \times 100\ (\%)$$

Figure 1.4 shows the magnitude of the point dose errors vs. composite field gamma passing rates in mock program. The gamma passing rates with 2%/2 mm criteria ranged much broader than those with 3%/3 mm even though there was no significant difference in point dose error. Data points of C-shape hard cases were the most broadly dispersed while data points of prostate cases were the most closely confined. There was no stringent correlation between the magnitude of point dose errors and gamma passing rates for composite field measurements.

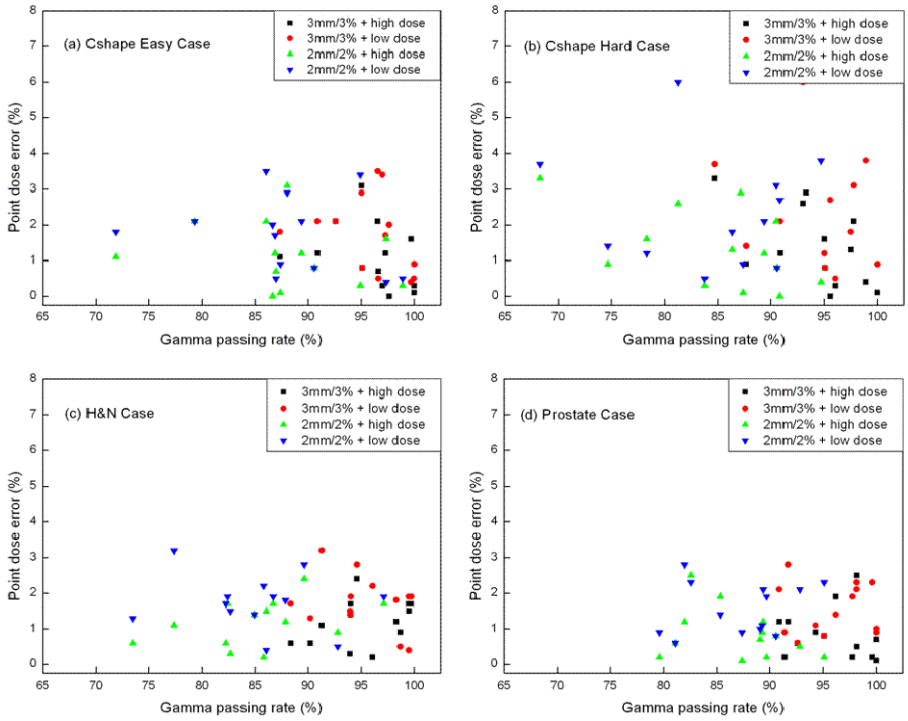

Figure 1.4. Magnitude of the point dose errors (low dose and high dose) versus composite filed gamma passing rates (2%/2 mm and 3%/3 mm) for each test: (a) C-shape easy, (b) C-shape hard, (c) H&N, and (d) prostate in the mock program.

Table 1.12. Per-field measurement: averaged percentage of points passing gamma criteria of 2%/2 mm and 3%/3 mm, with associated confidence limits in clinic program

Criteria	Test	Brain	H&N	Abdomen	Prostate	Overall
2%/2 mm	Mean (%)	93.9	91.0	91.1	92.9	92.4
	SD (%)	7.1	8.8	8.2	6.3	7.4
	CL (%)	80.1	73.7	75.1	80.5	77.9
3%/3 mm	Mean (%)	98.1	97.1	97.8	97.5	97.5
	SD (%)	3.3	4.4	3.1	2.9	3.4
	CL (%)	91.7	88.5	91.7	91.7	90.8
	N	14	22	10	45	91

Abbreviations: SD and CL stand for standard deviation and confidence limit.

Table 1.13. Per-field measurement: local and total averaged percentage of points passing gamma criteria of 2%/2 mm and 3%/3 mm, with associated confidence limits in clinic program

	Institution	A	B	C	D	E	F	G	H
Criteria	Device	Matri-XX	EBT2	Map-check	2D array	Matri-XX	Matri-XX	Matri-XX	Map-check
2%/ 2 mm	Mean (%)	92.0	94.9	88.7	96.8	91.3	91.4	97.1	88.8
	SD (%)	12.5	15.3	23.3	12.2	21.2	23.2	19.8	11.9
	LCL (%)	87.5	84.7	76.7	87.8	78.8	76.8	80.2	71.3
3%/ 3 mm	Mean (%)	96.1	98.7	95.6	99.1	97.2	98.1	99.3	96.3
	SD (%)	1.0	0.9	3.4	1.7	3.1	3.0	4.6	4.2
	LCL (%)	94.0	97.0	88.9	95.7	91.2	92.3	90.3	88.1
	N	9	7	10	12	20	10	11	12

Abbreviations: SD and LCL stand for standard deviation and local confidence limit.

Table 1.14. Composite-field measurement: averaged percentage of points passing gamma criteria of 2%/2 mm and 3%/3 mm, with associated confidence limits in clinic program

Group	Criteria	Test	Brain	H&N	Abdomen	Prostate	Overall
LINAC	2%/2 mm	Mean (%)	96.0	88.8	92.9	93.9	92.9
		SD (%)	4.0	10.1	11.9	4.0	7.4
		CL (%)	88.0	69.0	69.6	86.0	78.4
	3%/3 mm	Mean (%)	99.4	98.2	99.2	97.8	98.3
		SD (%)	0.8	2.0	1.6	2.1	1.9
		CL (%)	97.7	94.3	96.0	93.7	94.5
		N	14	22	10	45	61
TOMO	2%/2 mm	Mean (%)	84.1	92.1	85.2	87.5	89.6
		SD (%)	8.2	6.6	7.6	6.8	7.4
		CL (%)	68.1	79.2	70.4	74.2	75.1
	3%/3 mm	Mean (%)	95.8	97.6	96.4	97.0	97.2
		SD (%)	3.0	2.3	1.8	1.9	2.3
		CL (%)	89.9	93.0	92.9	93.3	92.7
		N	6	38	8	12	64

Abbreviations: SD and CL stand for standard deviation and confidence limit.

Table 1.15. Composite-field measurement: local and total averaged percentage of points passing gamma criteria of 2%/2 mm and 3%/3 mm, with associated confidence limits in clinic program

	Institution	A	B	C	D	E	F	G	H	I	J	K
Criteria	Device	Matr-ixx	EBT2	Map-check	2D-array	Matr-ixx	Matr-ixx	Matr-ixx	Map-check	EBT2	EBT2	EBT2
2%/ 2 mm	Mean (%)	91.6	92.3	93.0	96.8	93.5	96.7	88.5	90.0	95.4	91.4	83.3
	SD (%)	2.3	9.4	4.3	2.6	3.9	3.6	16.3	5.7	3.2	5.9	6.4
	LCL (%)	87.0	73.9	84.7	91.7	85.9	89.7	56.7	78.7	89.2	79.8	70.8
3%/ 3 mm	Mean (%)	96.5	98.8	98.7	99.8	97.1	99.6	99.1	97.5	98.6	97.5	95.7
	SD (%)	1.2	0.8	1.4	0.4	2.3	0.9	1.8	2.2	1.5	2.3	2.0
	LCL (%)	94.1	97.3	96.0	99.1	92.7	97.9	95.5	93.2	95.5	93.0	91.7
	N	9	7	10	12	20	10	11	12	20	20	24

Abbreviations: SD and LCL stand for standard deviation and local confidence limit.

DISCUSSION

In this study, the first multi-institutional study in Korea was performed to investigate and suggest tolerance levels for the patient-specific IMRT QA by adopting the confidence limit concept presented in AAPM TG-119 report [21]. The concept of confidence limit suggested by Venselaar *et al.* expressed with the mean value and the SD multiplied by 1.5 has been evaluated as a useful tool to quantify the dose accuracy of photon beam calculations of 3D treatment planning [47]. This study was different from AAPM TG-119 in terms of using different structures for mock program, which reflected the anatomy of average Korean, potentially possible to affect IMRT plans.

As shown in results, a concept of confidence limit for the point dose measurements was appropriate metrics because of a small standard deviation and a large number of samples. However, for planar dose gamma evaluations with 2%/2 mm criteria, it was not appropriate metrics since the data showed a large local deviation and the number of samples was small. With 3%/3 mm criteria, it was also difficult to define the confidence limit because of a small number of samples. There was no significant difference in the tolerance levels of point dose measurements between LINAC and TOMO groups and our tolerance levels agreed with those of AAPM and ESTRO guidelines. It could be concluded that the level of patient-specific QA for IMRT in Korea seems to meet the standard of international guidelines.

The result can be used as a reference data for other institutions in Korea when they evaluate their IMRT commissioning and patient-specific QA results.

Moreover, it could be used for the purpose of domestic audit for patient-specific IMRT QA to compare institutional local values.

In this study, the results demonstrated that the tolerance level of patient-specific QA in Korea met properly the international standards based on the concept of confidence limit. However, even though the tolerance level in Korea met the international standards, the meaning of the patient-specific QA results, *i.e.* sensitivity to detect errors in treatment plan is unclear. Furthermore, conventional patient-specific IMRT QA procedures have been applied to VMAT QA in spite of the differences from IMRT in terms of operation and planning mechanism. The ability of different patient-specific QA methods to detect errors in treatment plans of IMRT and VMAT was investigated in Chapter 2.

CHAPTER 2

A Study on Various Patient-specific Quality Assurances Concerning the Sensitivity to Detect Errors in IMRT and VMAT

INTRODUCTION

VMAT has attracted increasingly attention since it has been applied to the clinic because of its superior delivery efficiency over IMRT characterized by a shorter treatment time and lower MU [20, 48-50]. Clinical studies to assess the benefits of VMAT over IMRT have been actively investigated since the advent of VMAT [17, 19, 20, 48, 49, 51-54]. VMAT is a rotational delivery technique, and is distinguished from IMRT in that the modulated beam is achieved using variable dose rates, gantry rotation speeds, and MLC positions [50]. An arc of VMAT is approximated by control points in TPS, representing a set of static fields characterized by their gantry angle, MLC aperture, and weight of MU. The high complexity of VMAT requires precise and suitable QA procedures before the treatment of patients. Conventional patient-specific IMRT QA procedures such as point dose measurements using ion chamber or 2D planar dose measurements have been applied to VMAT QA. A variety of studies have been reported the results of VMAT QA using various methods of QA [55-60]. These studies addressed that the results of VMAT QA were acceptable for the appropriate treatment of patients.

The patient-specific QA result is a quantitative analysis of a measurement with tolerance level used to determine whether the plan is appropriate for patient treatment or not. To determine the appropriate tolerance level of IMRT QA, AAPM TG 119 carried out a multi-institutional investigation to assess the overall accuracy of planning and delivery and produced quantitative confidence limits as a baseline of expectation values for IMRT commissioning [36]. Similarly, the tolerance level in Korea has been investigated by multi-

institutional study involving 12 radiation therapy institutions in Chapter 1. In addition, several studies recommended the specific values to determine the tolerable level for IMRT QA. These tolerance levels generally have been applied to VMAT QA. Mancuso *et al.* compared the action levels of IMRT and VMAT QA using the geometry of structural set provided by AAPM TG 119 and reported that no statistically significant differences were observed [61]. Recent studies demonstrated that 3D dose-volumetric QA would be more appropriate for IMRT QA than 2D QA or the point dose measurement. Furthermore, it was demonstrated that 3D QA was more meaningful for VMAT than IMRT [62, 63]. Stasi *et al.* showed the correlation between gamma index and DVH information in patient-specific IMRT QA [64]. They concluded that no correlation was observed between QA results using different methodologies. Betzel *et al.* investigated whether VMAT was more susceptible to delivery uncertainties than dynamic IMRT [65]. Based on the recently performed studies, an appropriate QA system for VMAT and IMRT seems to be uncertain and disputable.

Even though variety of studies have been actively performed about QA systems for IMRT or VMAT, no study has been yet performed to investigate the sensitivities and correlations of QA methodologies covering a point dose measurement, 2D gamma evaluation and 3D volume dose analysis according to delivery technique and the degree of intensity modulation.

The aim of this study was to investigate the sensitivities of various QA methods to detect errors in IMRT and VMAT plans. Furthermore, the correlations of various QA methods were investigated in order to find out the

substitutability between different QA methods. The IMRT QA and VMAT QA results for patients with prostate and H&N cancer were analyzed comprehensively. Various procedures of QA were performed with (1) the point dose measurements using the ion chamber, (2) 2D dose distribution measurements of axial plane using radiochromic film (EBT2®, ISP, Wayne, NJ, USA), (3) 2D measurements of coronal plane using ion chamber array, MatriXX® (IBA dosimetry, Schwarzenbruck, Germany), and (4) the 3D volume dose distribution measurements using COMPASS® (IBA dosimetry, Schwarzenbruck, Germany). The measured data were analyzed statistically by grouping according to the delivery technique and the degree of intensity modulation. The concept of confidence limit was applied to investigate the tolerance levels. Finally, the correlations among results from different QA methodologies were investigated one another.

MATERIALS AND METHODS

1. Treatment planning

Ten of each prostate and H&N patients (total 20 patients) were selected for both IMRT and VMAT planning and delivery. A total of 40 treatment plans were generated for this study (20 IMRT and 20 VMAT plans). Each of the dynamic IMRT plans had 8 coplanar fields. Each of the VMAT plans had 2 coplanar full arcs. Both the IMRT and VMAT plans were generated using a 6 MV photon beam. The prescription dose was 44 Gy in 22 fractions. The optimization and dose calculation were done using Eclipse version 8.9.17 (Varian Medical Systems, Palo Alto, CA, USA). The optimization algorithm used for IMRT was the dose volume optimizer (DVO) while that of VMAT was the progressive resolution optimizer (PRO2). After the optimization, doses for both techniques were calculated by the analytical anisotropic algorithm (AAA) with a calculation gird of 0.25 cm. All plans were delivered using a Clinac iX (Varian Medical Systems, Palo Alto, CA, USA) equipped with the millennium MLC.

2. Patient-specific quality assurances

2-1. Point dose measurement

A point dose was measured with a 0.125 cc ion chamber (Semiflex ion chamber, PTW, Freiburg, Germany) in a custom-made cylindrical phantom made of acrylic. The dimensions of the phantom were 265 mm in length and 180 mm in diameter. There were two holes for insertion of an ion chamber. The phantom was composed of two pieces to allow insertion of a film. The two pieces of the phantom could be tightened using a lever in order to reduce the air gap between

the film and phantom. To eliminate the daily variation of machine output, a known dose was delivered and the ratio of chamber reading to the dose was acquired in unit of nC/cGy before DQA measurements. This ratio was applied to the results of IMRT and VMAT DQA.

2-2. Two-dimensional measurement of axial plane using film

Radiochromic EBT2 film was used to measure the axial 2D planar dose distributions for IMRT and VMAT DQA. The film was placed between the two pieces of a custom-made cylindrical phantom, where the isocenter was located at the center of the phantom. Film dosimetry carefully followed the self-developing procedure described in the ISP white paper [66]. Two batches of film were separately used for the measurements of total 40 plans. To avoid the inter-batch response variation of EBT films, which was known to be approximately less than 1%, the films from each batch-numbered packet were used for calibration [67]. The films were scanned after 20 hours of irradiation on a flatbed scanner (Epson 10000XL, Epson Canada Ltd., Toronto, Ontario) using 48 bit color mode (i.e., RGB mode) and practical spatial resolution of 75 dpi (i.e., approximately 0.2952 mm per pixel). The dual channel method of red and blue correction was applied for calibration [68, 69]. The calibration curve was acquired in the range of 0 to 350 cGy. The measured values of optical density were converted into to a dose map using in-house software written in C++. The calculated and measured dose distributions were compared using Verisoft 3.1 image software (PTW, Freiburg, Germany). The region of interest (ROI) was defined as a rectangle of 12×10 cm^2 and the threshold dose was set to be 5% of the maximum dose.

2-3. Two-dimensional measurement of coronal plane using MatriXX®

The MatriXX® (IBA dosimetry, Schwarzenbruck, Germany) was used to measure coronal plane doses. It is a 2D array consisting of 1020 ion chambers in a 32×32 grid covering a 24.4×24.4 cm2 active area. The device with back-scatter material (solid water phantom of 10 cm thickness) was fixed orthogonal to the beam direction at 100 cm source to surface distance (SSD) using a gantry fixture. This setup eliminates the angular dependence of the device and allows dose calculation at a gantry angle of 0° in TPS. Prior to the measurement, the device was warmed up for at least 30 minutes and pre-irradiated with 10 Gy for stabilization purpose, [70, 71] and then the background signals were compensated for. OmniPro IMRT software (IBA dosimetry, Schwarzenbruck, Germany) was used to acquire the dose distributions and to evaluate the measurement data.

2-4. Three-dimensional verification using COMPASS®

COMPASS® (IBA dosimetry, Schwarzenbruck, Germany) is a dosimetric tool which makes it possible to evaluate the volumetric dose by 3D dose reconstruction based on the fluence measurement and patient CT image set. The dose engine implemented in COMPASS® software uses the collapsed cone convolution/superposition (CCC) algorithm [72-74]. The sampling time for the measurement was 250 ms. An external angle sensor measured the actual gantry positions, which were reflected in 3D dose reconstruction. The MatriXX® was fixed orthogonal to the beam direction with 2 cm build-up material using the gantry fixture at 100 cm SSD. The accuracy of the detector setup was

thoroughly evaluated. The angle sensor was calibrated prior to the measurement. Once the whole treatment plan was delivered, the dose was reconstructed on the CT images of patient using measured fluences. The 3D DQA using COMPASS® generated the volumetric dose information for various structures such as body (totally irradiated volume), planning target volume (PTV), and organ at risks (OARs).

3. Verification of plan delivery

Two different log files were created by the MLC and machine controller, respectively. The MLC dynalog file is a record of the actual dose fraction or gantry angle versus actual MLC leaf positions. The machine dynalog file contains the treatment setup information and dynamic beam statistics (standard deviation of dose and dose-position). The file also contains information about the planned cumulative dose versus gantry angle, and delivered cumulative dose versus gantry angle. These data were used to verify whether the systematic error of dynamic delivery existed or not. The accuracy of the MLC leaf position for both IMRT and VMAT were verified in the dynalog file viewer, a utility program capable of analyzing the dynalog file data. The accuracy of MLC positioning during measurements was calculated as a root mean square (RMS) value of each leaf deviation. The accuracies of gantry angles and cumulative MUs during VMAT delivery were statistically evaluated in mean differences and standard deviation between the planned and actual values at each control point.

4. Data analysis and Statistics

The degree of modulation of each plan was described as total MU divided by the prescription dose.

The measured point dose was compared to the planned dose calculated by TPS, which was taken as the mean value inside the volume of ion chamber. The percent dose difference (%Diff) was computed as follows:

$$\%\text{Diff} = \frac{Measured\ dose - planned\ dose}{planned\ dose} \times 100\ (\%)$$

Two-dimensional dose distributions acquired by film and MatriXX® measurements were analyzed using gamma method with criteria of 2%/2 mm and 3%/3 mm.

The 2D dose distributions acquired by film and MatriXX® were analyzed using the gamma method with criteria of 2%/2 mm and 3%/3 mm. The measured 3D dose data with COMPASS® were also compared to the planned 3D doses of various structures using gamma evaluation with criteria of 2%/2 mm and 3%/3 mm. The structures for prostate cases were body, PTV, bladder, and rectum while those for H&N cases were body, PTV, brainstem, spinal cord, parotids, and esophagus.

The results were grouped by the delivery technique (IMRT and VMAT) and the tumor site (H&N and prostate) for the analysis. The Shapiro-Wilk test was performed to determine whether the data set of each group was well-modeled by a normal distribution or not [75]. In order to assess the statistical significance

of the difference between two groups, t-test was used if both of the groups followed the normal distribution, otherwise, Wilcoxon rank-sum test was used [76]. The Spearman coefficient was applied to evaluate the correlation among the DQA results of various dosimetric methodologies and different criteria [77, 78]. The confidence coefficient of 1.96 was applied for the normally distributed groups, otherwise, the confidence coefficient in the t-distribution table with two tails was applied depending on the number of samples within 95% confidence.

5. Arc discretization in VMAT

The VMAT plan of 90° arc was converted into a static arc (SA) plan with 97 control points. One control point (CP) in VMAT was regarded as one field of the SA plan. For this purpose, we developed an in-house program written in MATLAB (The MathWorks, Inc., Natick, MA). In the original VMAT plan, an arc is modeled by a sequence of CPs, defined on an aperture basis, equally spaced roughly every 2 degree at the end of the optimization process. Each MU at CP is in the form of cumulative meterset values in the DICOM file that indicate at what MU the machine should be at a certain gantry angle. These values are computed by setting the first one to 0 and then adding MUs for each CP. A *CP MU (C1 through C5)* is MUs to be delivered between start CP and end CP. Dose calculation in TPS was performed with the same gantry angle resolution as the progressive CP number. However, when doses in SA were computed, *Dose MU (D1 through D6)* was used instead of CP MU. All parameters of one CP were one static field's parameters so that a single SA plan had 97 static fields. The VMAT and SA plans coincided in terms of planning parameters. Figure 2.1 shows the SA plan generation and MU configuration.

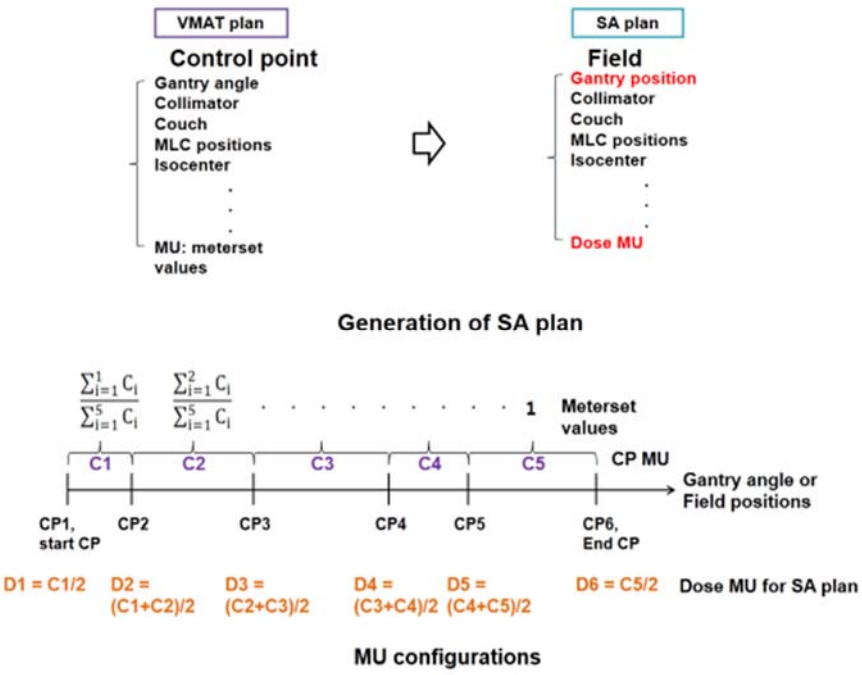

Figure 2.1. SA plan generation and MU configuration.

RESULTS

Degree of intensity modulation of IMRT and VMAT plans

The degrees of intensity modulation of each plan are listed in Table 2.1. The averaged value for IMRT plans in the prostate group was 4.2 MU/cGy ranging from 3.4 MU/cGy to 4.8 MU/cGy while that of the H&N group was 8.4 MU/cGy ranging from 5.3 MU/cGy to 11.6 MU/cGy, showing a large difference between the prostate and H&N groups. However, the averaged values for VMAT plans of prostate and H&N groups were 2.8 MU/cGy and 2.5 MU/cGy, respectively. The difference between the two groups in VMAT plans was negligible.

Table 2.1. The degree of intensity modulation calculated by total MU divided by prescription dose of 200 cGy

Patient	Prostate		H&N	
	IMRT (MU/cGy)	VMAT (MU/cGy)	IMRT (MU/cGy)	VMAT (MU/cGy)
1	4.1	2.1	8.2	2.3
2	3.7	3.2	8.6	2.0
3	4.8	2.9	6.6	2.9
4	3.4	3.4	8.5	2.4
5	3.8	3.2	11.6	2.0
6	4.2	2.7	10.9	2.8
7	4.5	1.9	8.6	2.0
8	4.6	2.8	7.9	2.9
9	4.3	2.8	7.6	3.2
10	4.2	2.9	5.3	2.3
Avg.	4.2	2.8	8.4	2.5

Abbreviations: H&N, IMRT, VMAT, MU, and Avg. stand for head and neck, intensity modulated radiation therapy, volumetric modulated arc therapy, monitor unit, and averaged value.

Verification of IMRT and VMAT delivery

The two dynalog files generated during IMRT and VMAT delivery were analyzed to detect the systematic error of machine performance. The deviation of MLC leaf position was less than 0.25 cm for all delivered fields. The mean and maximum values of RMS for MLC leaf motion were 0.05 cm and 0.08 cm, respectively. Figure 2.1 shows the deviations of the MU and gantry angle for a sample VMAT delivery. The largest MU deviation was observed at the starting control point in each arc of VMAT delivery, which ranged from -0.06 MU to 0.08 MU. Similarly, the largest gantry angle deviation was observed at the starting control point, ranging from -0.8° to 0.9°. These deviations were minuscule, and thus it could be concluded that there was no noticeable systematic error in the machine performance during measurements.

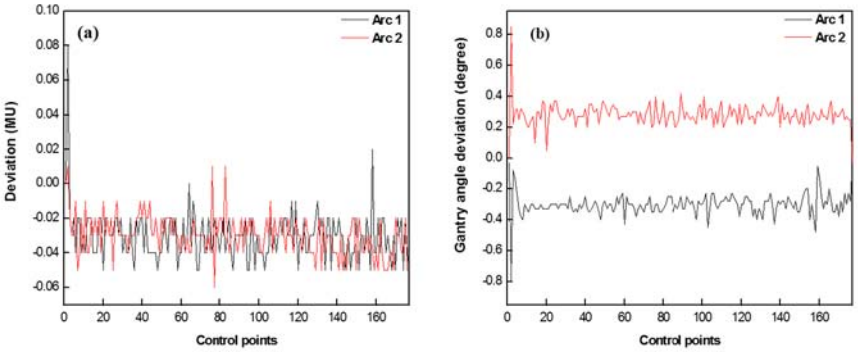

Figure 2.2. Deviations of MU (a) and gantry angle (b) acquired from Clinic dynalog file during VMAT delivery.

Point dose measurements

Data grouped by IMRT, VMAT, prostate, and H&N followed the normal distribution as confirmed by the Shapiro-Wilk test ($p > 0.05$). Thus, the confidence coefficient of 1.96 was applied for all groups. The averaged values of differences between planned and measured for IMRT and VMAT groups were -0.7 ± 1.2% and 0.4 ± 0.9%, respectively. The corresponding CLs were 3.0% and 2.1% for IMRT and VMAT groups, respectively. This difference between the two groups was statistically significant ($p < 0.05$). The averaged values of differences for prostate and H&N groups were 0.0 ± 0.5% and -0.4 ± 1.6%, respectively. The corresponding CLs for prostate and H&N groups were 1.0% and 3.4%, respectively. However, the difference between the two groups was not statistically significant ($p = 0.30$). The averaged difference of all tests was -0.2 ± 1.2% and the corresponding CL was 2.5%. Table 2.2 shows a summary of the point dose measurement results.

Table 2.2. The summary of point dose measurements

Group	N	Mean (%)	SD (%)	SW (*p*-value)	Significance (*p*-value)	CL (%)
IMRT	20	-0.7	1.2	0.05	< 0.05*	3.0
VMAT	20	0.4	0.9	0.74		2.1
Prostate	20	0.0	0.5	0.17	0.30*	1.0
H&N	20	-0.4	1.6	0.97		3.4
Total	40	-0.2	1.2	0.06	-	2.5

Abbreviations: N, SD, SW, CL, IMRT, VMAT, and H&N stand for the number of analyzed plans, standard deviation, Shapiro-Wilk test, confidence limit only available when the data followed normal distribution, intensity modulated radiation therapy, volumetric modulated arc therapy, and head and neck. *: t-test was used.

Two-dimensional measurements

Measurements of axial dose planes with film

Table 2.3 shows a summary of the averaged gamma passing rate of film measurements. Data were grouped by four categories and analyzed with two gamma criteria of 2%/2 mm and 3%/3 mm. Only the prostate data followed the normal distribution per the Shapiro-Wilk test. Therefore, the Wilcoxon rank-test was adopted to evaluate the statistical significance of differences among the grouped data. The confidence coefficient of 1.96 was applied to the group following the normal distribution. The confidence coefficient of 2.093 for 20 samples was applied to the groups that did not follow the normal distribution and 2.023 for 40 samples was applied to the total data.

The averaged gamma passing rates for IMRT and VMAT groups with the criteria of 2%/2 mm were 86.8 ± 8.8% and 82.8 ± 12.6%, respectively, and no statistically significant difference was observed ($p > 0.05$). The averaged gamma passing rates for prostate and H&N groups with the criteria of 2%/2 mm were 89.1 ± 5.5% and 80.5 ± 13.0%, respectively. The difference between the two groups was statistically significant ($p < 0.05$). The total results did not follow the normal distribution ($p < 0.05$) and the averaged gamma passing rate was 84.8 ± 10.8%. With 3%/3 mm criteria, the averaged gamma passing rates for IMRT and VMAT groups were 94.6 ± 5.3% and 91.1 ± 10.2%, respectively, and no statistically significant difference was observed ($p > 0.05$). To compare, the averaged gamma passing rates for prostate and H&N groups with 3%/3 mm criteria were 96.5 ± 2.1% and 89.3 ± 10.4%, respectively and the difference

was also statistically significant ($p < 0.05$). The total results did not follow the normal distribution ($p < 0.05$) with averaged gamma passing rate of $92.9 \pm 8.2\%$.

Table 2.3. The summary of 2D measurements in axial dose plane with film

Group	Criteria	N	Mean (%)	SD (%)	SW (*p*-value)	Significance test (*p*-value)	CL (%)
IMRT	2%/ 2 mm	20	86.8	8.5	0.01	0.37**	69.0†
VMAT		20	82.8	12.6	0.01		56.5†
Prostate		20	89.1	5.5	0.77	< 0.05**	78.4
H&N		20	80.5	13.0	0.03		53.3†
Total		40	84.8	10.8	0.00	-	63.0†
IMRT	3%/ 3 mm	20	94.6	5.3	0.00	0.20**	83.6†
VMAT		20	91.1	10.2	0.00		69.6†
Prostate		20	96.5	2.1	0.20	<0.05**	92.4
H&N		20	89.3	10.4	0.00		67.5†
Total		40	92.9	8.2	0.00	-	76.3†

Abbreviations: N, SD, SW, CL, IMRT, VMAT, and H&N stand for the number of analyzed plans, standard deviation, Shapiro-Wilk test, confidence limit only available when the data followed normal distribution, intensity modulated radiation therapy, volumetric modulated arc therapy, and head and neck. **: Wilcoxon rank-test was used, † was based on t-distribution.

Measurements of coronal dose planes with MatriXX®

Table 2.4 shows a summary of the averaged gamma passing rates with MatriXX®. Data was grouped into four categories and analyzed with two gamma criteria of 2%/2 mm and 3%/3 mm. Data grouped by prostate and H&N with 2%/2 mm criteria and H&N data with 3%/3 mm criteria followed the normal distribution per the Shapiro-Wilk test. The Wilcoxon rank-test was adopted to evaluate the statistical significance of differences except for the comparison between prostate and H&N groups with 2%/2 mm criteria. For the latter case, the t-test was adopted to evaluate the statistical significance of differences. The confidence coefficient of 1.96 was applied for the group following the normal distribution. The confidence coefficient of 2.093 for 20 samples was applied for the groups that did not follow the normal distribution and 2.023 for 40 samples was applied for the total data.

With criteria of 2%/2 mm, the averaged gamma passing rates for IMRT and VMAT groups were 78.0 ± 17.4% and 85.4 ± 11.2%, respectively. The difference between the two groups was statistically significant ($p < 0.05$).The values for prostate and H&N groups with 2%/2 mm were 92.6 ± 2.3% and 70.8 ± 14.2%, respectively and the difference was statistically significant ($p < 0.05$). The total results did not follow the normal distribution ($p < 0.05$) and the averaged gamma passing rate was 81.7 ± 14.9%.

With criteria of 3%/3 mm, the averaged gamma passing rates for IMRT and VMAT groups were 88.7 ± 10.8% and 94.4 ± 6.5%, respectively. The difference between the two groups was statistically significant ($p < 0.05$). The averaged gamma passing rate for prostate and H&N groups with criteria of

3%/3 mm were 97.3 ± 1.5% and 85.8 ± 10.3%, respectively, and the difference was statistically significant ($p < 0.05$). The total results did not follow the normal distribution ($p < 0.05$) with an averaged gamma passing rate of 91.5 ± 9.3%.

Table 2.4. The summary of the 2D measurements in coronal dose plane with MatriXX®

Group	Criteria	N	Mean (%)	SD (%)	SW (*p*-value)	Significance test (*p*-value)	CL (%)
IMRT	2%/ 2 mm	20	78.0	17.4	0.01	< 0.05**	41.6†
VMAT		20	85.4	11.2	0.00		62.0†
Prostate		20	92.6	2.3	0.45	< 0.05*	88.1
H&N		20	70.8	14.2	0.83		43.1
Total		40	81.7	14.9	0.00	-	51.6†
IMRT	3%/ 3 mm	20	88.7	10.8	0.00	< 0.05**	66.1†
VMAT		20	94.4	6.5	0.00		80.8†
Prostate		20	97.3	1.5	0.02	< 0.05**	94.2†
H&N		20	85.8	10.3	0.22		65.7
Total		40	91.5	9.3	0.00	-	72.7†

Abbreviations: N, SD, SW, CL, IMRT, VMAT, and H&N stand for the number of analyzed plans, standard deviation, Shapiro-Wilk test, confidence limit only available when the data followed normal distribution, intensity modulated radiation therapy, volumetric modulated arc therapy, and head and neck. *: t-test was used, **: Wilcoxon rank-test was used, † was based on t-distribution.

Three-dimensional volume dose analysis

The results of the 3D DQA were used to analyze the difference between VMAT and IMRT deliveries within the structures of interest within each treatment site group. Moreover, the gamma passing rates of both PTV and body structures were used as representative value for the analysis of four categories with two gamma criteria of 2%/2 mm and 3%/3 mm.

Prostate cases

The results of the prostate data are summarized in Table 2.5. Per the Shapiro-Wilk test, the normal distribution was given to the following: the body and PTV data for VMAT group with the criteria of 2%/2 mm, the body, PTV, bladder, and rectum data for IMRT group with the criteria of 2%/2 mm, the body, PTV, and rectum data in VMAT group with criteria of 3%/3 mm, and the body and rectum data in IMRT group with criteria of 3%/3 mm. The t-test was adopted to evaluate statistical significance of differences when comparing IMRT versus VMAT data for body and PTV with the criteria of 2%/2 mm and IMRT versus VMAT data for body and rectum with the criteria of 3%/3 mm. For the rest of the cases, the Wilcoxon rank-test was adopted for the same purpose. The confidence coefficient of 1.96 was applied for the group following the normal distribution. The confidence coefficient of 2.262 for 10 samples was applied for the groups that did not follow the normal distribution.

The averaged gamma passing rates of IMRT and VMAT data for PTV with the criteria of 2%/2 mm were 97.6 ± 1.4% and 79.0 ± 6.8%, respectively and showed statistically significant differences ($p < 0.05$). With the criteria of 3%/3 mm, IMRT and VMAT data for PTV and rectum showed statistically

significant differences ($p < 0.05$). The averaged gamma passing rate of IMRT and VMAT groups for PTV were 99.8 ± 0.2% and 97.2 ± 2.1%, respectively.

Table 2.5. The summary of 3D volume dose analysis for various structures in prostate group.

Volume	Group	Criteria	N	Mean (%)	SD (%)	SW (*p*-value)	Significance test (*p*-value)	CL (%)
Body	IMRT	2%/ 2 mm	10	98.5	0.3	0.28	0.93*	97.7
	VMAT		10	98.5	0.5	0.81		97.5
PTV	IMRT		10	97.6	1.4	0.25	< 0.05*	94.9
	VMAT		10	79.0	6.8	0.56		65.7
Bladder	IMRT		10	98.6	0.9	0.19	< 0.05**	96.8
	VMAT		10	97.0	2.7	0.02		90.9†
Rectum	IMRT		10	88.9	4.5	0.27	0.85**	80.1
	VMAT		10	89.7	2.4	0.01		84.3†
Body	IMRT	3%/ 3 mm	10	99.4	0.2	0.41	0.11*	99.0
	VMAT		10	99.6	0.2	0.23		99.2
PTV	IMRT		10	99.8	0.2	0.01	< 0.05**	99.3†
	VMAT		10	97.2	2.1	0.38		93.1
Bladder	IMRT		10	99.9	0.2	0.01	0.85**	99.4†
	VMAT		10	99.8	0.3	0.01		99.1†
Rectum	IMRT		10	98.7	1.2	0.17	0.05*	96.4
	VMAT		10	97.3	1.5	0.30		94.5

Abbreviations: N, SD, SW, CL, IMRT, VMAT, and H&N stand for the number of analyzed plans, standard deviation, Shapiro-Wilk test, confidence limit only available when the data followed normal distribution, intensity modulated radiation therapy, volumetric modulated arc therapy, and head and neck. *: t-test was used, **: Wilcoxon rank-test was used, † was based on t-distribution.

Head and neck cases

The results of H&N data were summarized in Table 2.6. Per the Shapiro-Wilk test, the normal distribution was given to the following: the body and PTV in VMAT group with criteria of 2%/2 mm, the PTV, spinal cord, right and left parotid in IMRT group with criteria of 2%/2 mm, the body and PTV in VMAT group with criteria of 3%/3 mm, and right and left parotid in IMRT group with criteria of 3%/3 mm. The t-test was adopted to evaluate the statistical significance of differences only when comparing IMRT versus VMAT results of PTV with criteria of 2%/2 mm. For the rest of the cases the Wilcoxon rank-test was adopted for the same purpose. The confidence coefficient of 1.96 was applied for the group following the normal distribution. The confidence coefficient of 2.262 for 10 samples was applied for the groups that did not follow the normal distribution.

With criteria of 2%/2 mm and 3%/3 mm, for comparison between IMRT and VMAT, only the results of body, PTV, and spinal cord showed a statistically significant difference ($p < 0.05$). The averaged gamma passing rates of the IMRT and VMAT groups for PTV with the criteria of 2%/2 mm were 85.7 ± 8.2% and 58.6 ± 12.0%, respectively. With criteria of 3%/3 mm, the averaged gamma passing rates of the IMRT and VMAT groups for PTV were 94.9 ± 5.2% and 83.8 ± 8.6%, respectively.

Table 2.6. The summary of 3D volume dose analysis for various structures in head and neck group

Volume	Group	Criteria	N	Mean (%)	SD (%)	SW (*p*-value)	Significa-nce test (*p*-value)	CL (%)
Body	IMRT	2%/ 2 mm	10	98.8	0.9	0.01	< 0.05**	96.8†
	VMAT		10	98.3	1.1	0.81		96.2
PTV	IMRT		10	85.7	8.2	0.21	< 0.05*	69.7
	VMAT		10	58.6	12.0	0.46		35.2
Brainstem	IMRT		10	97.1	6.9	0.00	0.69**	81.5†
	VMAT		10	94.9	15.7	0.00		59.4†
Spinal cord	IMRT		10	93.0	4.9	0.30	< 0.05**	83.4
	VMAT		10	99.2	1.3	0.00		96.3†
Rt parotid	IMRT		10	96.4	3.1	0.10	0.85**	90.4
	VMAT		10	94.8	6.4	0.01		80.3†
Lt parotid	IMRT		10	97.9	1.4	0.31	0.77**	95.2
	VMAT		10	97.0	3.8	0.00		88.4†
Esophagus	IMRT		10	94.8	6.6	0.02	0.91**	79.9†
	VMAT		10	93.2	10.9	0.00		68.5†
Body	IMRT	3%/ 3 mm	10	99.8	0.2	0.00	< 0.05**	99.3†
	VMAT		10	99.6	0.3	0.30		99.0
PTV	IMRT		10	94.9	5.2	0.04	< 0.05**	83.1†
	VMAT		10	83.8	8.6	0.51		67.0
Brainstem	IMRT		10	99.7	1.0	0.00	1.00**	97.4†
	VMAT		10	97.7	7.3	0.00		81.2†
Spinal cord	IMRT		10	99.3	1.0	0.01	< 0.05**	97.0†
	VMAT		10	100.0	0.1	0.00		99.8†
Rt parotid	IMRT		10	99.4	0.7	0.06	0.82**	98.0
	VMAT		10	98.9	1.8	0.00		94.8†
Lt parotid	IMRT		10	99.8	0.1	0.14	0.91**	99.6
	VMAT		10	99.7	0.4	0.01		98.8†
Esophagus	IMRT		10	99.8	0.4	0.00	0.31**	98.9†
	VMAT		10	98.5	2.9	0.00		91.9†

Abbreviations: N, SD, SW, CL, IMRT, VMAT, and H&N stand for the number of analyzed plans, standard deviation, Shapiro-Wilk test, confidence limit only available when the data followed normal distribution, intensity modulated radiation therapy, volumetric modulated arc therapy, and head and neck. *: t-test was used, **: Wilcoxon rank-test was used, † was based on t-distribution.

Comparison between groups

The results of PTV and body volume were analyzed by comparing the IMRT group to the VMAT group and the prostate group to the H&N group. The results are summarized in Table 2.7. Since the body and PTV in VMAT group with criteria of 2%/2 mm, the body in both prostate and H&N groups with criteria of 2%/2 mm, PTV in H&N group with both criteria of 2%/2 mm and 3%/3 mm, body in both IMRT and VMAT groups with criteria of 3%/3 mm, body in prostate group with criteria of 3%/3 mm, and body in total group with criteria of 3%/3 mm followed the normal distribution according to Shapiro-Wilk test, the t-test was adopted to evaluate statistical significances of differences when comparing prostate vs. H&N results of body with criteria of 2%/2 mm and IMRT vs. VMAT results of body with criteria of 3%/3 mm. For the rest of the cases the Wilcoxon rank-test was adopted. The confidence coefficient of 1.96 was applied for the group following the normal distribution. The confidence coefficient of 2.093 for 20 samples was applied for the groups that did not follow the normal distribution and 2.023 for 40 samples was applied for the total data.

The comparison between the prostate and H&N groups for PTV regardless of criteria showed statistically significant differences ($p < 0.05$). When comparing the IMRT group with the VMAT group for PTV, regardless of criteria, the differences were statistically significant ($p < 0.05$).

With criteria of 2%/2 mm, the averaged gamma passing rate of body in the prostate and H&N groups were 98.5 ± 0.4% and 98.5 ± 1.0%, respectively.

Those values with criteria of 3%/3 mm were 99.5 ± 0.2% in the prostate group and 99.7 ± 0.3% in the H&N group.

The averaged gamma passing rates of PTV in the IMRT and VMAT groups with criteria of 2%/2 mm were 91.7 ± 8.4% and 68.8 ± 14.1%, respectively, showing large differences with statistical significance. Those values were 88.3 ± 10.7% in the prostate group and 72.2 ± 17.1% in the H&N group which also showed large differences. With criteria of 3%/3 mm, the averaged gamma passing rates of PTV in the IMRT and VMAT groups were 97.3 ± 4.4% and 90.5 ± 9.2%, respectively. Even though the difference was reduced relative to the difference with 2%/2 mm criteria, it was still considerable. The averaged gamma passing rates of PTV in the prostate and H&N groups were 98.5 ± 2.0% and 89.3 ± 9.0%, respectively, showing large differences.

Table 2.7. The summary of 3D volume dose analyses for body and PTV structures

Volume	Group	Criteria	N	Mean (%)	SD (%)	SW (*p*-value)	Significance test (*p*-value)	CL (%)
Body	IMRT	2%/ 2 mm	20	98.6	0.7	0.01	< 0.05**	97.1†
	VMAT		20	98.4	0.8	0.63		96.8
PTV	IMRT		20	91.7	8.4	0.00	< 0.05**	74.1†
	VMAT		20	68.8	14.1	0.28		41.3
Body	Prostate		20	98.5	0.4	0.46	0.75*	97.7
	H&N		20	98.5	1.0	0.06		96.6
PTV	Prostate		20	88.3	10.7	0.00	< 0.05**	65.9†
	H&N		20	72.2	17.1	0.20		38.9
Body	Total		40	98.5	0.8	0.02	-	96.9†
PTV	Total		40	80.2	16.3	0.00	-	47.2†
Body	IMRT	3%/ 3 mm	20	99.6	0.3	0.14	0.69*	
	VMAT			99.0				
			20	99.6	0.2	0.36		99.2
PTV	IMRT							
	VMAT		20	97.3	4.4	0.00	< 0.05**	88.1†
			20	90.5	9.2	0.01		71.2†
Body	Prostate							
	H&N		20	99.5	0.2	0.41	< 0.05**	99.1
			20	99.7	0.3	0.00		99.1†
PTV	Prostate							
	H&N		20	98.5	2.0	0.00	< 0.05**	94.3†
Body	Total		20	89.3	9.0	0.06		71.8
PTV	total							

Abbreviations: N, SD, SW, CL, IMRT, VMAT, and H&N stand for the number of analyzed plans, standard deviation, Shapiro-Wilk test, confidence limit only available when the data followed normal distribution, intensity modulated radiation therapy, volumetric modulated arc therapy, and head and neck. *: t-test was used, **: Wilcoxon rank-test was used, † was based on t-distribution.

Statistical correlation among different protocols of quality assurance

The results of statistical correlations represented by *r*-values among results from different QA protocols such as point dose measurement, 2D gamma evaluation, and 3D gamma evaluation are shown in Table 2.8 along with *p*-values. The *r*-values of point dose difference versus other QA methodologies such as MatriXX®, Film, and 3D gamma passing rate of Body and PTV with COMPASS® were negative values of -0.33 ($p < 0.05$), -0.32 ($p = 0.06$), -0.36 ($p < 0.05$), and -0.31 ($p = 0.05$) respectively. This means point dose difference increased with decreasing of the gamma passing rate. The *r*-values of 2D gamma passing rate with MatriXX® vs. film and 2D gamma passing rate with MatriXX® vs. 3D gamma passing rate of PTV with COMPASS® were positive values of 0.43 ($p < 0.05$) and 0.42 ($p < 0.05$), respectively. The *r*-values of 2D gamma passing rate with film vs. 3D gamma passing rate of body and PTV with COMPASS® were also positive values of 0.03 ($p = 0.86$) and 0.50 ($p < 0.05$), respectively. This means that they were contradictory to each other. However, the absolute values of *r*-values were always less than 0.8 indicating weak correlations. For the other cases, no correlations were observed with one another. Figure 2.3 shows the correlations between the point dose difference and 2D or 3D gamma passing rate with criteria of 3%/3 mm.

Table 2.8. Spearman correlation coefficient (*r*-value) among different QA protocols (criteria of 3%/3 mm for 2D and 3D measurement)

Measurement		2D gamma evaluation (MatriXX®)	2D gamma evaluation (Film)	3D gamma evaluation of Body	3D gamma evaluation of PTV
Point dose difference	*r*-value	-0.33	-0.30	-0.36	-0.31
	p-value	< 0.05	0.06	< 0.05	0.05
	N	40	40	40	40
2D gamma evaluation (MatriXX®)	*r*-value		0.43	-0.05	0.42
	p-value		< 0.05	0.78	< 0.05
	N		40	40	40
2D gamma evaluation (Film)	*r*-value			0.03	0.50
	p-value			0.86	< 0.05
	N			40	40
3D gamma evaluation of body	*r*-value				-0.02
	p-value				0.90
	N				40

Abbreviations: PTV and N stand for planning target volume and the number of analyzed plans.

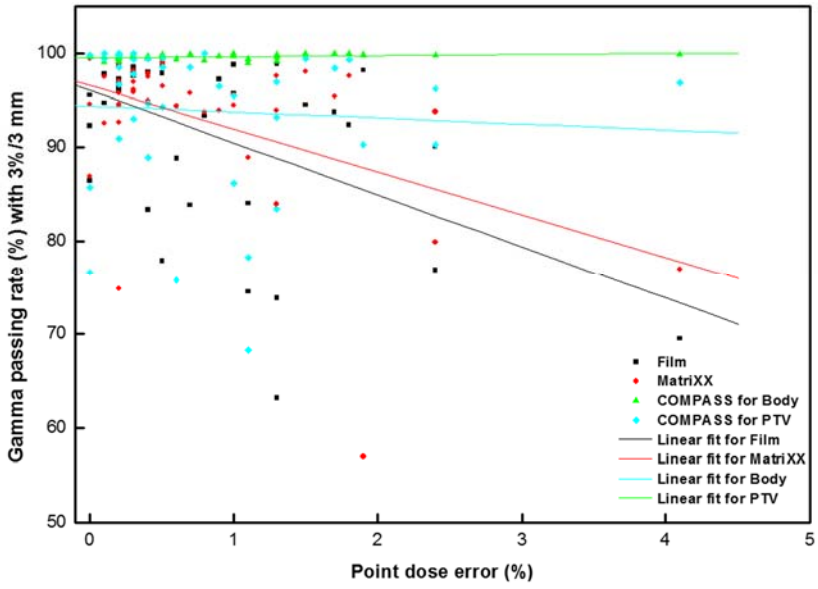

Figure 2.3. Spearman correlation coefficient (*r*-value) for the results of point dose difference vs. gamma passing rate with criteria of 3%/3 mm using film, MatriXX®, and COMPASS®.

Statistical correlation between gamma passing rate with criteria of 2%/2 mm and 3%/3 mm

The results of statistical correlations between gamma passing rate with criteria of 2%/2 mm and 3%/3 mm from measurements using film, MatriXX®, and COMPASS® are shown in Table 2.9. The *r*-values were 0.94 for the results with MatriXX®, 0.93 for the film, 0.77 for Body with COMPASS®, and 0.93 for PTV with COMPASS®. All of these values are above 0.8, indicating strong correlation with *p*-values less than 0.05. Figure 3 shows the positive correlation between the gamma passing rate with criteria of 2%/2 mm and 3%/3 mm.

Table 2.9. Spearman correlation coefficient (*r*-value) between gamma passing rate with criteria of 2%/2 mm and 3%/3 mm

Analysis		MatriXX® 2%/2 mm	Film 2%/2 mm	COMPASS® 2%/2 mm (Body)	COMPASS 2%/2 mm (PTV)
MatriXX® 3%/3 mm	*r*-value	0.94			
	p-value	< 0.05			
	N	40			
Film 3%/3 mm	*r*-value		0.93		
	p-value		< 0.05		
	N		40		
COMPASS® 3%/3 mm (Body)	*r*-value			0.77	
	p-value			< 0.05	
	N			40	
COMPASS® 3%/3 mm (PTV)	*r*-value				0.93
	p-value				< 0.05
	N				40

Abbreviations: PTV and N stand for planning target volume and the number of analyzed plans.

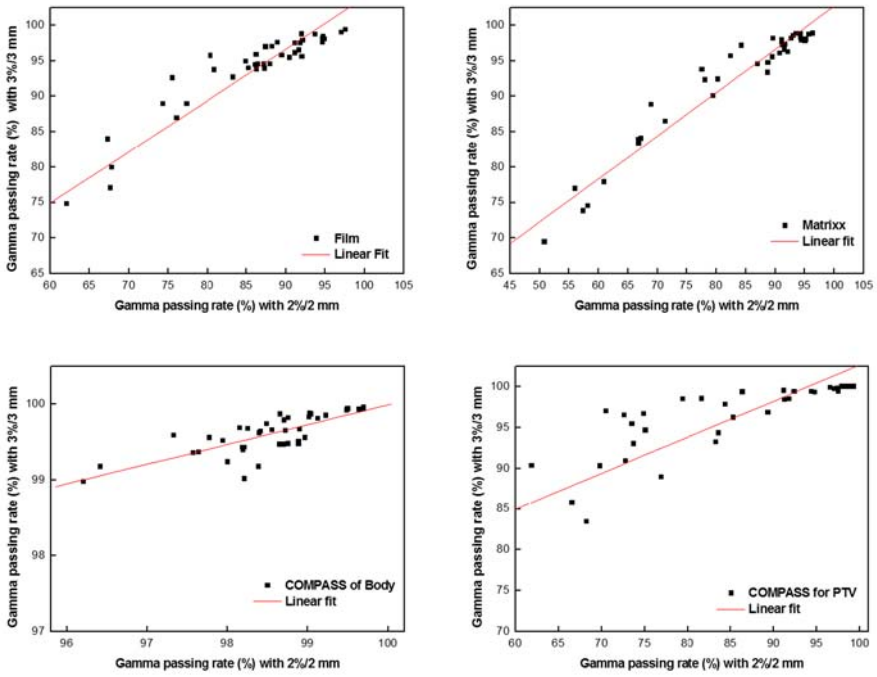

Figure 2.4. Correlation between gamma passing rates with different criteria.

The reduction of dose discrepancy in PTV of VMAT and SA QA

The dose discrepancy of PTV between VMAT and SA plans were evaluated using by gamma test with 3%/3 mm criteria and dose difference at 95% and 5% volume of PTV, respectively. The gamma passing rate of QA results were 92.1% and 96.8% for VMAT and SA QA, respectively. The dose differences at D_{95} were 2.61% and 0.97% for VMAT and SA QA, respectively. The dose differences at D_5 were 2.71% and 0.04% for VMAT and SA QA, respectively. The QA results of SA plan showed the less dose discrepancy with gamma passing rate and dose difference in PTV than VMAT QA.

DISCUSSION

In this study, comprehensive patient-specific QA using different QA methodologies and different delivery techniques were performed. For the reliable collection of data, the linac performance was verified with log files recorded during beam delivery. With thorough verification of machine, we assumed that the delivery of IMRT and VMAT was faultless. Based on this assumption, the QA results were analyzed by grouping according to the degree of intensity modulation, QA methodology and beam delivery technique. The correlations among the groups were evaluated in order to investigate the sensitivity of each QA methodology to detect errors in VMAT and IMRT. The higher degree of intensity modulation is believed to be associated with higher possibility to be failed in QA [79-87]. The degree of intensity modulation represented by MU/cGy of H&N group was 2 times higher than that of prostate group in IMRT while the tendency was slightly reversed in VMAT. However, the difference in VMAT was negligible with a value of about 10%. In addition, the fundamental intensity modulation mechanism of VMAT is different from that of IMRT, which is acquired by changing MU and MLC apertures per control points. Therefore, the degree of intensity modulation of VMAT could be determined considering various factors such as the degree of MLC movement, the frequency of small field usage and dose rate fluctuation rather than single value of MU/cGy. Consequently, it might be assumed that the overall degree of modulation of H&N group was higher than that of prostate group regardless of delivery technique. For that reason, the QA results of H&N

group might be expected to be worse than that of prostate group. Based on this assumption, the sensitivities of various QA methods to detect the errors in IMRT and VMAT were evaluated in this study.

In the results of point dose measurements, the averaged QA results and the CLs were acceptable according to the international guidelines [34, 36]. However, no considerable differences were observed between IMRT and VMAT group as well as between prostate and H&N group.

In the results of 2D gamma evaluations using film or MatriXX®, the results of H&N group were worse than those of prostate group with statistical significance. Two-dimensional gamma evaluation appears to be more capable of detecting the errors in beam delivery of VMAT and IMRT. Some results of 2D gamma evaluation seemed to be not acceptable for the treatment of patients considering the tolerance level of international guidelines [36]. This was due to the more strict 2D QA procedure of this study than routine QA procedure used in clinic. The dose distributions were compared with absolute values rather than relative values and the normalization point was consistent in applying at the point of CAX in this study. Furthermore, the daily fluctuation of machine output was not considered.

The results of 3D gamma evaluation for PTV in H&N group were much worse than those of prostate group with statistical significances regardless of criteria. The 3D results of body in H&N group were similar with those in prostate group. Since the body was the largest structure, partial volume of body was irradiated. With this reason, it might not be observed noticeable differences for body not only between IMRT and VMAT group but also between prostate and H&N

group. In the results of 3D gamma evaluation for PTV, the differences of passing rate between IMRT group and VMAT group were more enhanced with the criteria of 2%/2 mm than with the criteria of 3%/3 mm. The cause of low passing rate of PTV in 3D QA for VMAT was partially originated in the intrinsic nature of optimization algorithm of TPS known as arc discretization. Besides of that, Bhagwat *et al*. studied the MLC uncertainty in VMAT and this could be advocate the deviation observed in our study [88]. Consequently, it was demonstrated that the point dose measurement was not sensitive enough to be used as a QA method for IMRT and VMAT, similar with previous studies [89, 90]. The 2D or 3D gamma evaluation seems to be capable of detecting errors in IMRT and VMAT delivery. Even though both 2D and 3D gamma evaluation could be appropriate methods of IMRT or VMAT QA, 3D gamma evaluation provides more information than 2D QA. 3D gamma evaluation using COMPASS® is capable of evaluating the differences between calculated and measured dose distribution of each organs separately. The discrepancies in PTV structure of VMAT were only detected by the 3D QA with COMPASS®. Therefore, 3D QA seems to be an appropriate QA method for IMRT and VMAT. Stasi *et al*. recently studied the correlation between gamma index and DVH in the patient-specific IMRT QA [64]. They concluded that the results of various QA methods were not correlated strongly one another. Furthermore, they asserted that the published acceptance criteria have disputable predictive power for patient-specific IMRT QA. Similarly, no correlations were observed between different QA methods in our study. Therefore, point dose measurement or 2D gamma evaluation could not be alternatives to the 3D gamma evaluation

regarded as the most appropriate patient-specific QA method for IMRT and VMAT. The correlations between criteria of 2%/2 mm and 3%/3 mm were also investigated. Even though strong correlation was observed between them, criteria of 2%/2 mm appeared more sensitive than 3%/3 mm criteria.

In this study, various patient-specific QA methods for IMRT and VMAT technique such as point dose measurement, 2D and 3D gamma evaluation were performed. As an appropriate QA method for IMRT or VMAT, gamma evaluation of 3D volume dose seems to be ideal. The point dose measurement or 2D gamma evaluation could not be alternatives to the 3D gamma evaluation since no correlation was observed between point dose or 2D gamma evaluation and 3D gamma evaluation. Even in 2D gamma evaluation in coronal and axial plane shows no correlation and different results. Through the 3D QA, the discrepancy between calculated and measured dose distribution in PTV structure were found. In addition, through 3D QA for VMAT, we found that the discrepancy between calculated and measured dose distribution in PTV. By further study, some part of this discrepancy was due to the arc discretization of TPS when calculating the dose distribution of VMAT.

GENERAL DISCUSSION

The methodologies of QA or tolerance levels for IMRT and VMAT to deliver appropriate prescription dose and to spare normal tissue were rigorously studied and recommended by groups of experts such as AAPM, ASTRO or ESTRO [35, 36, 39]. Patient-specific QA methods widely-adopted in clinic are point dose measurement, gamma evaluation of 2D dose distribution and 3D volume dose analysis [24, 28, 59, 91-97]. Our tolerance levels of point dose and 2D dose measurements by multi-institutional study agreed with those of AAPM and ESTRO guidelines. It could be concluded that the level of patient-specific QA for IMRT in Korea seems to meet the standard of international guidelines. The result can be used as a reference data for other institutions in Korea when they evaluate their IMRT commissioning and patient-specific QA results. However, even though a reasonable tolerance level was derived by a multi-institutional study, it is uncertain that which method is suitable for detecting errors in IMRT and VMAT. Through the study of various patient-specific QA methods for IMRT and VMAT technique, 3D QA was the most sensitive and appropriate QA method for IMRT and VMAT. The sensitive QA tools to provide 3D information are need to assess the sophisticated machine performance and to evaluate the clinical outcome with the results of QA. Further work by other 3D verification metrics is required to assess the dose discrepancy in VMAT and the multi-institutional study is required to have confidence in results in this study.

REFERENCES

1. E. Martin, C. Deville, F. Bonnetain, M. Bosset, G. Crehange, G. Truc, S. Naudy, et al., "Intensity-modulated radiation therapy in head and neck cancer: prescribed dose, clinical challenges and results," Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology 85, 392-398 (2007).
2. M. Grant, "The effect of nursing consultation on anxiety, side effects, and self-care of patients receiving radiation therapy," Oncology nursing forum 17, 31-36; discussion 36-38 (1990).
3. V. Gregoire, and T. R. Mackie, "State of the art on dose prescription, reporting and recording in Intensity-Modulated Radiation Therapy (ICRU report No. 83)," Cancer radiotherapy 15, 555-559 (2011).
4. R. Yaparpalvi, L. Hong, D. Mah, J. Shen, S. Mutyala, M. Spierer, et al., "ICRU reference dose in an era of intensity-modulated radiation therapy clinical trials: correlation with planning target volume mean dose and suitability for intensity-modulated radiation therapy dose prescription," Radiotherapy and oncology 89, 347-352 (2008).
5. J. E. Bayouth, and S. M. Morrill, "MLC dosimetric characteristics for small field and IMRT applications," Medical physics 30, 2545-2552 (2003).
6. J. E. Bayouth, D. Wendt, and S. M. Morrill, "MLC quality assurance techniques for IMRT applications," Medical physics 30, 743-750 (2003).

7. D. Georg, and B. Kroupa, "Pre-clinical evaluation of an inverse planning module for segmental MLC based IMRT delivery," Physics in medicine and biology 47, N303-314 (2002).

8. K. Otto, and B. G. Clark, "Enhancement of IMRT delivery through MLC rotation," Physics in medicine and biology 47, 3997-4017 (2002).

9. C. B. Saw, R. A. Siochi, K. M. Ayyangar, W. Zhen, and C. A. Enke, "Leaf sequencing techniques for MLC-based IMRT," Med Dosim 26, 199-204 (2001).

10. K. M. Ayyangar, P. S. Nizin, C. B. Saw, D. Gearheart, B. Shen, and C. A. Enke, "Independent dose calculations for the corvus MLC IMRT," Med Dosim 26, 135-141 (2001).

11. C. B. Saw, K. M. Ayyangar, W. Zhen, R. B. Thompson, and C. A. Enke, "Commissioning and quality assurance for MLC-based IMRT," Med Dosim 26, 125-133 (2001).

12. C. B. Saw, K. M. Ayyangar, and C. A. Enke, "MLC-based IMRT-Part II," Med Dosim 26, 111-112 (2001).

13. P. M. Bruch, and W. Zhen, "Viewpoints to consider in the development of MLC and IMRT programs," Med Dosim 26, 91-94 (2001).

14. J. Deng, T. Pawlicki, Y. Chen, J. Li, S. B. Jiang, and C. M. Ma, "The MLC tongue-and-groove effect on IMRT dose distributions," Physics in medicine and biology 46, 1039-1060 (2001).

15. S. Webb, "A new concept of multileaf collimator (the shuttling MLC)--an interpreter for high-efficiency IMRT," Physics in medicine and biology 45, 3343-3358 (2000).

16. N. Hardcastle, W. A. Tome, K. Foo, A. Miller, M. Carolan, and P. Metcalfe, "Comparison of prostate IMRT and VMAT biologically optimised treatment plans," Med Dosim 36, 292-298 (2011).

17. C. Sale, and P. Moloney, "Dose comparisons for conformal, IMRT and VMAT prostate plans," Journal of medical imaging and radiation oncology 55, 611-621 (2011).

18. C. L. Tsai, J. K. Wu, H. L. Chao, Y. C. Tsai, and J. C. H. Cheng, "Treatment and Dosimetric Advantages between Vmat, Imrt, and Helical Tomotherapy in Prostate Cancer," Med Dosim 36, 264-271 (2011).

19. R. W. Kopp, M. Duff, F. Catalfamo, D. Shah, M. Rajecki, and K. Ahmad, "VMAT vs. 7-field-IMRT: assessing the dosimetric parameters of prostate cancer treatment with a 292-patient sample," Med Dosim 36, 365-372 (2011).

20. D. Wolff, F. Stieler, G. Welzel, F. Lorenz, Y. Abo-Madyan, S. Mai, C. Herskind, et al., "Volumetric modulated arc therapy (VMAT) vs. serial tomotherapy, step-and-shoot IMRT and 3D-conformal RT for treatment of prostate cancer," Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology 93, 226-233 (2009).

21. G. A. Ezzell, J. W. Burmeister, N. Dogan, T. J. LoSasso, J. G. Mechalakos, D. Mihailidis, et al., "IMRT commissioning: multiple institution planning and dosimetry comparisons, a report from AAPM Task Group 119," Medical physics 36, 5359-5373 (2009).

22. G. Yan, C. Liu, T. A. Simon, L. C. Peng, C. Fox, and J. G. Li, "On the sensitivity of patient-specific IMRT QA to MLC positioning errors," Journal of applied clinical medical physics / American College of Medical Physics 10, 2915 (2009).

23. G. Heilemann, B. Poppe, and W. Laub, "On the sensitivity of common gamma-index evaluation methods to MLC misalignments in Rapidarc quality assurance," Medical physics 40, 031702 (2013).

24. W. van Elmpt, S. Nijsten, B. Mijnheer, A. Dekker, and P. Lambin, "The next step in patient-specific QA: 3D dose verification of conformal and intensity-modulated RT based on EPID dosimetry and Monte Carlo dose calculations," Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology 86, 86-92 (2008).

25. V. Chandraraj, S. Stathakis, R. Manickam, C. Esquivel, S. S. Supe, and N. Papanikolaou, "Comparison of four commercial devices for RapidArc and sliding window IMRT QA," Journal of applied clinical medical physics / American College of Medical Physics 12, 3367 (2011).

26. D. W. Bailey, L. Kumaraswamy, and M. B. Podgorsak, "A fully electronic intensity-modulated radiation therapy quality assurance (IMRT QA) process implemented in a network comprised of independent treatment planning, record and verify, and delivery systems," Radiology and oncology 44, 124-130 (2010).

27. T. Pawlicki, S. Yoo, L. E. Court, S. K. McMillan, R. K. Rice, J. D. Russell, et al., "Process control analysis of IMRT QA: implications for clinical trials," Physics in medicine and biology 53, 5193-5205 (2008).

28. R. M. Howell, I. P. Smith, and C. S. Jarrio, "Establishing action levels for EPID-based QA for IMRT," Journal of applied clinical medical physics / American College of Medical Physics 9, 2721 (2008).

29. B. E. Nelms, and J. A. Simon, "A survey on planar IMRT QA analysis," Journal of applied clinical medical physics / American College of Medical Physics 8, 2448 (2007).

30. M. R. Zaini, G. J. Forest, and D. D. Loshek, "Image processing for IMRT QA dosimetry," Journal of applied clinical medical physics / American College of Medical Physics 6, 14-21 (2005).

31. J. M. Moran, J. Radawski, and B. A. Fraass, "A dose gradient analysis tool for IMRT QA," Journal of applied clinical medical physics / American College of Medical Physics 6, 62-73 (2005).

32. P. S. Basran, and M. K. Woo, "An analysis of tolerance levels in IMRT quality assurance procedures," Medical physics 35, 2300 (2008).

33. G. A. Ezzell, J. M. Galvin, D. Low, J. R. Palta, I. Rosen, M. B. Sharpe, et al., "Guidance document on delivery, treatment planning, and clinical implementation of IMRT: Report of the IMRT subcommittee of the AAPM radiation therapy committee," Medical physics 30, 2089 (2003).

34. J. R. Palta, J. A. Deye, G. S. Ibbott, J. A. Purdy, and M. M. Urie, "Credentialing of institutions for IMRT in clinical trials," Int J Radiat Oncol Biol Phys 59, 1257-1259; author reply 1259-1261 (2004).

35. C. H. Clark, V. N. Hansen, H. Chantler, C. Edwards, H. V. James, G. Webster, et al., "Dosimetry audit for a multi-centre IMRT head and neck trial," Radiother Oncol 93, 102-108 (2009).

36. G. A. Ezzell, J. W. Burmeister, N. Dogan, T. J. LoSasso, J. G. Mechalakos, D. Mihailidis, et al., "IMRT commissioning: Multiple institution planning and dosimetry comparisons, a report from AAPM Task Group 119," Medical physics 36, 5359 (2009).

37. G. Budgell, J. Berresford, M. Trainer, E. Bradshaw, P. Sharpe, and P. Williams, "A national dosimetric audit of IMRT," Radiother Oncol 99, 246-252 (2011).

38. R. I. MacKay, J. Staffurth, A. Poynter, D. Routsis, and B. Radiotherapy Development, "UK guidelines for the safe delivery of intensity-modulated radiotherapy," Clin Oncol (R Coll Radiol) 22, 629-635 (2010).

39. C. H. Clark, E. A. Miles, M. T. Urbano, S. A. Bhide, A. M. Bidmead, K. J. Harrington, et al., "Pre-trial quality assurance processes for an intensity-modulated radiation therapy (IMRT) trial: PARSPORT, a UK multicentre Phase III trial comparing conventional radiotherapy and parotid-sparing IMRT for locally advanced head and neck cancer," Br J Radiol 82, 585-594 (2009).

40. J. B. Chung, J. S. Kim, S. W. Ha, and S. J. Ye, "Statistical analysis of IMRT dosimetry quality assurance measurements for local delivery guideline," Radiat Oncol 6, 27 (2011).

41. B. Vanderstraeten, N. Reynaert, L. Paelinck, I. Madani, C. De Wagter, W. De Gersem, et al., "Accuracy of patient dose calculation for lung IMRT: A comparison of Monte Carlo, convolution/superposition, and pencil beam computations," Medical physics 33, 3149-3158 (2006).

42. J. C. L. Chow, M. K. K. Leung, and J. Van Dyk, "Variations of lung density and geometry on inhomogeneity correction algorithms: A Monte Carlo dosimetric evaluation," Medical physics 36, 3619 (2009).

43. A. Sen, and M. K. West, "Commissioning experience and quality assurance of helical tomotherapy machines," J Med Phys 34, 194-199 (2009).

44. A. Van Esch, J. Bohsung, P. Sorvari, M. Tenhunen, M. Paiusco, M. Iori, P. Engstrom, et al., "Acceptance tests and quality control (QC) procedures for the clinical implementation of intensity modulated radiotherapy (IMRT) using inverse planning and the sliding window technique: experience from five radiotherapy departments," Radiotherapy and Oncology 65, 53-70 (2002).

45. L. D. Wolfsberger, M. Wagar, P. Nitsch, M. S. Bhagwat, and P. Zygmanski, "Angular dose dependency of MatriXX TM and its calibration," Journal of Applied Clinical Medical Physics 11, 241-251 (2010).

46. J. W. Lee, J. B. Chung, D. H. Lee, J. H. Park, B. Y. Choe, T. S. Suh, et al., "Discrepancy of Intensity Modulation Radiation Therapy Dose Delivery due to the Dose-Dynamic Multi-Leaf Collimator Gravity Effect," Journal of the Korean Physical Society 53, 3436-3443 (2008).

47. J. Venselaar, H. Welleweerd, and B. Mijnheer, "Tolerances for the accuracy of photon beam dose calculations of treatment planning systems," Radiotherapy and Oncology 60, 191-201 (2001).

48. E. M. Quan, X. Li, Y. Li, X. Wang, R. J. Kudchadker, J. L. Johnson, et al., "A comprehensive comparison of IMRT and VMAT plan quality for prostate cancer treatment," International journal of radiation oncology, biology, physics 83, 1169-1178 (2012).

49. M. Pasler, H. Wirtz, and J. Lutterbach, "Impact of Gantry Rotation Time on Plan Quality and Dosimetric Verification-Volumetric Modulated Arc Therapy (VMAT) vs. Intensity Modulated Radiotherapy (IMRT)," Strahlentherapie Und Onkologie 187, 812-819 (2011).

50. K. Otto, "Volumetric modulated arc therapy: IMRT in a single gantry arc," Medical physics 35, 310 (2008).

51. T. F. Lee, P. J. Chao, H. M. Ting, S. H. Lo, Y. W. Wang, C. C. Tuan, et al., "Comparative analysis of SmartArc-based dual arc volumetric-modulated arc radiotherapy (VMAT) versus intensity-modulated radiotherapy (IMRT) for nasopharyngeal carcinoma," Journal of applied clinical medical physics / American College of Medical Physics 12, 3587 (2011).

52. M. Pasler, D. Georg, H. Wirtz, and J. Lutterbach, "Effect of Photon-Beam Energy on VMAT and IMRT Treatment Plan Quality and Dosimetric Accuracy for Advanced Prostate Cancer," Strahlentherapie Und Onkologie 187, 792-798 (2011).

53. X. Q. Jiang, T. Li, Y. M. Liu, L. Zhou, Y. Xu, X. J. Zhou, et al., "Planning analysis for locally advanced lung cancer: dosimetric and efficiency comparisons between intensity-modulated radiotherapy (IMRT), single-arc/partial-arc volumetric modulated arc therapy (SA/PA-VMAT)," Radiation Oncology 6 (2011).

54. C. L. Tsai, J. K. Wu, H. L. Chao, Y. C. Tsai, and J. C. Cheng, "Treatment and dosimetric advantages between VMAT, IMRT, and helical tomotherapy in prostate cancer," Med Dosim 36, 264-271 (2011).

55. D. Letourneau, M. Gulam, D. Yan, M. Oldham, and J. W. Wong, "Evaluation of a 2D diode array for IMRT quality assurance," Radiother Oncol 70, 199-206 (2004).

56. V. Feygelman, G. Zhang, C. Stevens, and B. E. Nelms, "Evaluation of a new VMAT QA device, or the "X" and "O" array geometries," Journal of applied clinical medical physics / American College of Medical Physics 12, 3346 (2011).

57. G. Li, Y. Zhang, X. Jiang, S. Bai, G. Peng, K. Wu, a et al., "Evaluation of the ArcCHECK QA system for IMRT and VMAT verification," Phys Med (2012).

58. J. Adamson, and Q. Wu, "Independent verification of gantry angle for pre-treatment VMAT QA using EPID," Physics in medicine and biology 57, 6587-6600 (2012).

59. C. Wu, K. E. Hosier, K. E. Beck, M. B. Radevic, J. Lehmann, H. H. Zhang, et al., "On using 3D gamma-analysis for IMRT and VMAT pretreatment plan QA," Medical physics 39, 3051-3059 (2012).

60. B. E. Nelms, D. Opp, J. Robinson, T. K. Wolf, G. Zhang, E. Moros, et al., "VMAT QA: measurement-guided 4D dose reconstruction on a patient," Medical physics 39, 4228-4238 (2012).

61. G. M. Mancuso, J. D. Fontenot, J. P. Gibbons, and B. C. Parker, "Comparison of action levels for patient-specific quality assurance of intensity modulated radiation therapy and volumetric modulated arc therapy treatments," Medical physics 39, 4378-4385 (2012).

62. J. J. Kruse, "On the insensitivity of single field planar dosimetry to IMRT inaccuracies," Medical physics 37, 2516 (2010).

63. B. E. Nelms, H. Zhen, and W. A. Tomé, "Per-beam, planar IMRT QA passing rates do not predict clinically relevant patient dose errors," Medical physics 38, 1037 (2011).

64. M. Stasi, S. Bresciani, A. Miranti, A. Maggio, V. Sapino, and P. Gabriele, "Pretreatment patient-specific IMRT quality assurance: a correlation study between gamma index and patient clinical dose volume histogram," Medical physics 39, 7626-7634 (2012).

65. G. T. Betzel, B. Y. Yi, Y. Niu, and C. X. Yu, "Is RapidArc more susceptible to delivery uncertainties than dynamic IMRT?," Medical physics 39, 5882-5890 (2012).

66. "Gafchromic® EBT2 self-developing film for radiotherapy dosimetry," ISP White Paper (2010).

67. L. J. van Battum, D. Hoffmans, H. Piersma, and S. Heukelom, "Accurate dosimetry with GafChromic EBT film of a 6 MV photon beam in water: what level is achievable?," Medical physics 35, 704-716 (2008).

68. R. R. Mayer, F. Ma, Y. Chen, R. I. Miller, A. Belard, J. McDonough, et al., "Enhanced dosimetry procedures and assessment for EBT2 radiochromic film," Medical physics 39, 2147-2155 (2012).

69. A. Micke, D. F. Lewis, and X. Yu, "Multichannel film dosimetry with nonuniformity correction," Medical physics 38, 2523-2534 (2011).

70. J. Herzen, M. Todorovic, F. Cremers, V. Platz, D. Albers, A. Bartels, et al., "Dosimetric evaluation of a 2D pixel ionization chamber for implementation in clinical routine," Physics in medicine and biology 52, 1197-1208 (2007).

71. Z. H. Han, S. K. Ng, M. S. Bhagwat, Y. Lyatskaya, and P. Zygmanski, "Evaluation of MatriXX for IMRT and VMAT dose verifications in peripheral dose regions," Medical physics 37, 3704-3714 (2010).

72. F. Hasenbalg, H. Neuenschwander, R. Mini, and E. J. Born, "Collapsed cone convolution and analytical anisotropic algorithm dose

calculations compared to VMC++ Monte Carlo simulations in clinical cases," Physics in medicine and biology 52, 3679-3691 (2007).

73. A. Ahnesjo, "Collapsed cone convolution of radiant energy for photon dose calculation in heterogeneous media," Medical physics 16, 577-592 (1989).

74. O. I. Calvo, A. N. Gutierrez, S. Stathakis, C. Esquivel, and N. Papanikolaou, "On the quantification of the dosimetric accuracy of collapsed cone convolution superposition (CCCS) algorithm for small lung volumes using IMRT," Journal of applied clinical medical physics / American College of Medical Physics 13, 3751 (2012).

75. S. S. Shapiro, and M. B. Wilk, "An Analysis of Variance Test for Normality (Complete Samples)," Biometrika 52, 591-& (1965).

76. P. D. Bridge, and S. S. Sawilowsky, "Increasing physicians' awareness of the impact of statistics on research outcomes: comparative power of the t-test and and Wilcoxon Rank-Sum test in small samples applied research," J Clin Epidemiol 52, 229-235 (1999).

77. C. Spearman, "The proof and measurement of association between two things," American Journal of Psychology 15, 72-101 (1904).

78. C. Spearman, "The proof and measurement of association between two things," International Journal of Epidemiology 39, 1137-1150 (2010).

79. E. Astreinidou, A. Bel, C. P. Raaijmakers, C. H. Terhaard, and J. J. Lagendijk, "Adequate margins for random setup uncertainties in head-and-neck IMRT," Int J Radiat Oncol Biol Phys 61, 938-944 (2005).

80. K. S. Chao, G. Ozyigit, B. N. Tran, M. Cengiz, J. F. Dempsey, and D. A. Low, "Patterns of failure in patients receiving definitive and postoperative IMRT for head-and-neck cancer," Int J Radiat Oncol Biol Phys 55, 312-321 (2003).

81. B. Dabaja, M. R. Salehpour, I. Rosen, S. Tung, W. H. Morrison, K. K. Ang, et al., "Intensity-modulated radiation therapy (IMRT) of cancers of the head and neck: comparison of split-field and whole-field techniques," Int J Radiat Oncol Biol Phys 63, 1000-1005 (2005).

82. E. K. Hansen, M. K. Bucci, J. M. Quivey, V. Weinberg, and P. Xia, "Repeat CT imaging and replanning during the course of IMRT for head-and-neck cancer," Int J Radiat Oncol Biol Phys 64, 355-362 (2006).

83. M. van Vulpen, C. Field, C. P. Raaijmakers, M. B. Parliament, C. H. Terhaard, M. A. MacKenzie, et al., "Comparing step-and-shoot IMRT with dynamic helical tomotherapy IMRT plans for head-and-neck cancer," Int J Radiat Oncol Biol Phys 62, 1535-1539 (2005).

84. K. A. Vineberg, A. Eisbruch, M. M. Coselmon, D. L. McShan, M. L. Kessler, and B. A. Fraass, "Is uniform target dose possible in IMRT plans in the head and neck?," Int J Radiat Oncol Biol Phys 52, 1159-1172 (2002).

85. M. T. Studenski, V. Bar-Ad, J. Siglin, D. Cognetti, J. Curry, M. Tuluc, et al., "Clinical experience transitioning from IMRT to VMAT for head and neck cancer," Med Dosim (2012).

86. A. Molineu, N. Hernandez, T. Nguyen, G. Ibbott, and D. Followill, "Credentialing results from IMRT irradiations of an anthropomorphic head and neck phantom," Medical physics 40, 022101 (2013).

87. N. Ploquin, I. Kay, A. Rangel-Baltazar, H. Lau, and P. Dunscombe, "A comparison of techniques for simulating set-up error and uncertainty in head and neck IMRT," Medical physics 33, 3213-3219 (2006).

88. M. S. Bhagwat, Z. Han, S. K. Ng, and P. Zygmanski, "An oscillating sweeping gap test for VMAT quality assurance," Physics in medicine and biology 55, 5029-5044 (2010).

89. Y. Watanabe, "Point dose calculations using an analytical pencil beam kernel for IMRT plan checking," Physics in medicine and biology 46, 1031-1038 (2001).

90. A. Kumar, G. Mukherjee, G. Yadav, V. Pandey, and K. Bhattacharya, "Optimized point dose measurement: An effective tool for QA in intensity-modulated radiotherapy," J Med Phys 32, 156-160 (2007).

91. M. Zhang, S. Qin, T. Chen, S. Kim, S. Jabbour, B. Haffty, et al., "A clinical objective IMRT QA method based on portal dosimetry and electronic portal imager device (EPID) measurement," Technology in cancer research & treatment 12, 145-150 (2013).

92. A. J. Olch, "Evaluation of the accuracy of 3DVH software estimates of dose to virtual ion chamber and film in composite IMRT QA," Medical physics 39, 81-86 (2012).

93. Y. Nakaguchi, F. Araki, S. Saiga, T. Kouno, M. Maruyama, K. Kakei, et al., "Comparison of dose accuracy between 2D array detectors for

pre-treatment IMRT QA," Nihon Hoshasen Gijutsu Gakkai zasshi 68, 443-452 (2012).

94. R. Boggula, M. Birkner, F. Lohr, V. Steil, F. Wenz, and H. Wertz, "Evaluation of a 2D detector array for patient-specific VMAT QA with different setups," Physics in medicine and biology 56, 7163-7177 (2011).

95. K. A. Johansson, P. Nilsson, B. Zackrisson, B. Ohlson, E. Kjellen, C. Mercke, et al., "The quality assurance process for the ARTSCAN head and neck study - a practical interactive approach for QA in 3DCRT and IMRT," Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology 87, 290-299 (2008).

96. K. M. Langen, S. L. Meeks, D. O. Poole, T. H. Wagner, T. R. Willoughby, O. A. Zeidan, et al., "Evaluation of a diode array for QA measurements on a helical tomotherapy unit," Medical physics 32, 3424-3430 (2005).

97. A. Kapulsky, G. Gejerman, and J. Hanley, "A clinical application of an automated phantom-film QA procedure for validation of IMRT treatment planning and delivery," Med Dosim 29, 279-284 (2004).

국문 초록

서론: 국내 다기관 공동연구를 통하여 세기조절 방사선치료의 환자 맞춤형 정도관리 허용오차를 조사하고, 다양한 선량검증방법을 통해서 세기조절방사선치료와 입체세기조절회전 방사선치료에서의 환자 맞춤형 정도관리의 민감도를 분석하여 허용오차를 조사한다.

방법: 다기관 공동연구는 국내 12 개 기관이 참여하였고, 선량정도관리는 이온전리함을 이용한 점선량 측정과 필름이나 2 차원 선량 검출기를 이용한 평면 선량 측정 방법을 이용하였다. Mock 프로그램과 임상 프로그램으로 구성되었다. 세기조절 방사선치료와 입체세기조절회전 방사선치료에서의 다양한 선량검증방법은 이온전리함을 이용한 점선량 측정, 필름을 이용한 시상면 선량 측정, 2 차원 이온전리함을 이용한 관상면 선량 측정과 COMPASS 를 이용한 3 차원 입체적 선량 측정을 포함한다. 두경부암 환자군, 전립선암 환자군, 세기조절 방사선치료 방법과 입체세기조절회전 방사선치료 방법 등 4 가지 그룹으로 나뉘어 분석되었다. 각 비교 그룹의 정규성, 상관관계와 유의성 검증을 실시하였다. 추가적으로 새로운 아크분할 치료계획을 만들어 PTV 에서의 선량 불일치를 3 차원 정도관리를 통해서 평가하였다.

결과: 다기관 공동연구의 환자 맞춤형 세기조절 방사선치료의 정도관리 허용오차는 고점선량 ±3%, 저선량 ±5~7%로 나타났으며, 평명선

량에 대한 허용오차 범위는 큰 표준편차와 적은 표본수로 결정하기 어려웠다. 그러나 계산된 허용오차 범위는 AAPM 에서 제시한 값과 일치하였다. 환자 맞춤형 정도관리 민감도에 대한 연구에서 점선량의 허용오차는 세기조절 방사선치료, 입체세기조절회전 방사선치료, 전립선암 그리고 두경부암에서 3.0%, 2.1%, 1.0% 그리고 3.4%로 나타났다. 두 가지 평면선량 측정방법에서 모두 치료방법 및 환자군의 비교에서 모두 유의한 차이를 보였다. 3 차원 입체적 선량 측정에서 또한 치료방법 및 환자군의 비교에서 모두 유의한 차이를 보였다. 모든 선량 검증 방법에 따른 결과의 상관관계는 없었다. 그리고 3 차원 정도관리에서만 입체세기조절회전 방사선치료에서 치료계획 표적 부위 선량 불일치를 감지 할 수 있었다. 아크분할 치료계획에서 정도관리 결과 입체세기조절회전 방사선치료의 치료계획 표적 부위의 감마테스트 결과 92.1%에서 96.8%로 증가하였고, 치료계획 표적 부위의 선량 불일치 정도가 95% 체적이 받는 선량이 2.61%에서 0.97%로 5% 체적이 받는 선량이 2.71%에서 0.04%로 현저하게 감소하였다.

결론: 국내 다기관 공동연구의 결과는 AAPM 과 ESTRO 의 다기관 공동연구와 동일한 결과를 얻게 되었다. 두 가지 치료방법의 차이는 점선량, 평면 선량과 3 차원 입체선량에서 유의하였으며, 환자군의 따른 차이는 점선량을 제외한 평면 선량과 3 차원 입체선량에서 유의하게 나타났다. 3 차원 입체선량 방법이 가장 민감도가 크게 나타났으며, 입체세기조절회전 방사선치료의 치료계획 표적 부위 선량 불일치를 감

지 할 수 있었다. 따라서 세기조절 방사선치료와 입체세기조절회전 방사선치료를 이용한 정확한 방사선치료의 평가를 위해서는 3차원 선량 정도관리가 필요하다.

주요어 : 허용오차, 민감도, 정도관리, 다기관 연구, 세기조절 방사선치료, 입체세기조절회전 방사선치료

학 번 : 2009-30597