UNIVERSIDAD D CORDOBA

TESIS DOCTORAL

para optar al grado de

Doctor

por la

Universidad de Córdoba

Especialidad: Biomedicina

con la mención de *Doctorado Internacional*

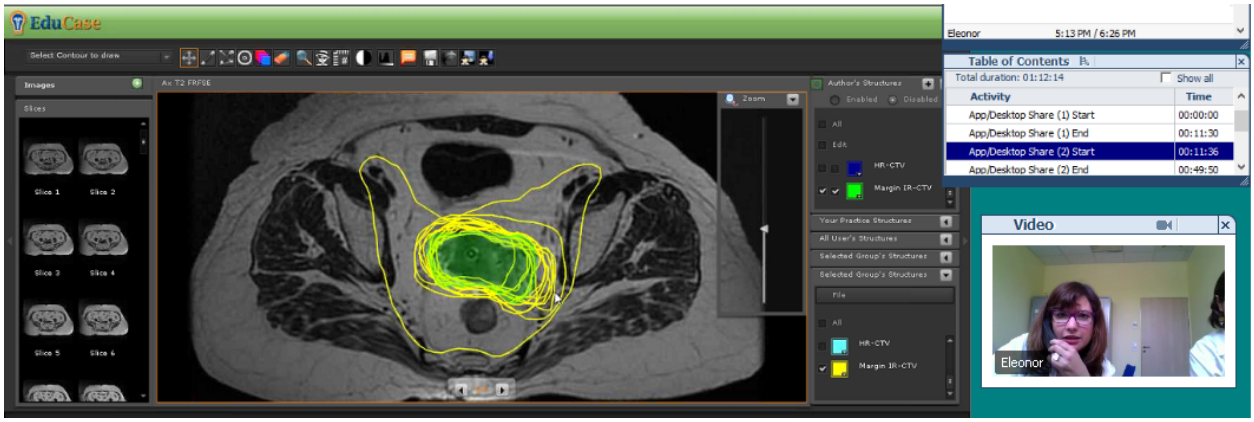
Presentada por

Eleonor Rivin del Campo

Diciembre de 2017

Radiotherapy quality control in cervical cancer

Control de calidad de la radioterapia del cáncer de cervix



Dirigida por

Pr. María Martínez-Paredes, Universidad de Córdoba, Córdoba, España
Dra. Amalia Palacios Eito, Hospital Universitario Reina Sofía, Córdoba, España
Pr. Eric Deutsch, Gustave Roussy Cancer Campus, Villejuif, Francia

TITULO: *Radiotherapy quality control in cervical cancer*

AUTOR: *Eleonor Rivin del Campo*

Campus de Rabanales
Ctra. Nacional IV, Km. 396 A
14071 Córdoba

https://www.uco.es/ucopress/index.php/es/
ucopress@uco.es

MARÍA MARTÍNEZ-PAREDES, Doctora en Medicina y Cirugía y Profesora Titular de Radiología y Medicina Física de la Universidad de Córdoba (España).

INFORMA:

Que Dña. Eleonor Rivin del Campo, ha desarrollado bajo mi dirección y supervisión el trabajo titulado : **Control de calidad de la radioterapia del cáncer de cervix (Radiotherapy Quality Control in Cervical Cancer).**

El desarrollo del trabajo ha evolucionado adecuadamente en el tiempo, generando una serie de comunicaciones y trabajos, y un artículo publicado en la revista Radiotherapy and Oncology, 124 (2017) 130-138. Por lo que, examinado y revisado, se da conformidad para su presentación y defensa ante el tribunal que le sea asignado para su juicio crítico y calificación, a fin de conseguir la obtención del Grado de Doctor en Medicina.

Y para que conste y surta efectos, firma el presente informe en Córdoba, a veinte de diciembre de dos mil diecisiete.

Fdo.: Profª. Dra. María Martínez-Paredes

UNIVERSIDAD DE CÓRDOBA

AMALIA PALACIOS EITO, Doctora en Medicina y Cirugía y Profesora Asociada del Departamento de Ciencias Sociosanitarias y Radiología y Medina Física de la Universidad de Córdoba (España).

INFORMA:

Que Dña. Eleonor Rivin del Campo, ha desarrollado bajo mi dirección el trabajo titulado: **Control de calidad de la radioterapia del Cáncer de Cervix (Radiotherapy Quality Control in Cervical Cancer).**

El desarrollo del trabajo ha seguido una sistemática de elaboración en tiempo, forma y contenido según el proyecto inicial. Se ha concluido con el mejor aprovechamiento, por lo que examinado y revisado, se da conformidad para que sea defendido y juzgado como TESIS para la obtención del Grado de Doctor en Medicina.

Y para que conste y surta efectos, firma el presente informe en Córdoba, a doce de diciembre de dos mil diecisiete.

Fdo.: Amalia Palacios Eito

TÍTULO DE LA TESIS: Control de calidad de la radioterapia del cáncer de cervix (Radiotherapy Quality Control in Cervical Cancer).

DOCTORANDO/A: Eleonor Rivin del Campo

ERIC DEUTSCH, Doctor in Medicine and full-Professor of Radiation Oncology in the South-Paris University (France).

I have directed and supervised Eleonor Rivin del Campo's PhD project: **Control de calidad de la radioterapia del cáncer de cervix (Radiotherapy Quality Control in Cervical Cancer).**

She has performed the task of Quality Control within a multicenter European FP-7 research project in cervical cancer which served as the frame for her PhD thesis for which I was a director.

The project has been performed in a timely fashion, meeting the deadlines imposed by the multicenter European research project. In parallel, the preparation and writing of her PhD thesis has also been done at an adequate pace. An article on this experience has been published in the journal Radiotherapy and Oncology (IF. 4.328), which holds the highest impact factor of all journals specialized in Radiation Oncology. Several communications (oral and poster), have been performed by the PhD student on the subject of this thesis, including an oral communication in the European Society for Radiotherapy and Oncology (ESTRO) Congress 33 held in 2014. Other oral communications have been presented by her in Córdoba, Spain and in Paris, France.

For all of these reasons, after having examined and reviewed this PhD thesis, I endorse its submission and its dissertation before the assigned jury for its evaluation with ends of obtaining the title of PhD.

Por todo ello, se autoriza la presentación de la tesis doctoral.

París, 15 de diciembre de 2017,

Firma del director

Fdo.: Pr. Eric Deutsch

TÍTULO DE LA TESIS:

Control de calidad de la radioterapia del cáncer de cérvix. (Radiotherapy quality control in cervical cancer)

DOCTORANDO/A:

ELEONOR RIVIN DEL CAMPO

INFORME RAZONADO DEL/DE LOS DIRECTOR/ES DE LA TESIS

(se hará mención a la evolución y desarrollo de la tesis, así como a trabajos y publicaciones derivados de la misma).

DÑA MARÍA MARTÍNEZ PAREDES Doctora en Medicina y Cirugía y Profesora Titular de Radiología y Medicina Física de la Universidad de Córdoba (España).

DÑA AMALIA PALACIOS EITO Doctora en Medicina y Cirugía y Profesora Asociada del Departamento de Ciencias Sociosanitarias y Radiología y Medicina Física de la Universidad de Córdoba (España).

D. ERIC DEUTSCH Doctor en Medicina y Profesor Titular de Oncología Radiotérapica y Radiobiología de la Universidad Paris-Sud (Francia).

HACEN CONSTAR

Que el trabajo titulado "**Control de calidad de la radioterapia del cáncer de cervix (Radiotherapy Quality Control in Cervical Cancer)** " ha sido realizado por Dña ELEONOR RIVIN DEL CAMPO bajo nuestra direccion dentro del Programa de Doctorado: **Biomedicina** y adscrito a la línea de investigación: **HORMONAS Y CÁNCER. NEUROPÉPTIDOS Y SUS RECEPTORES EN TUMORES HORMONO-DEPENDIENTES.**

Consideramos que esta tesis hace una aportación significativa al conocimiento de un tema relevante tal cual es el *control de calidad de la radioterapia en el cáncer de cérvix uterino*, validando una metodologia que permite su optimización en el marco de un ensayo multicéntrico.

Este trabajo ha conseguido un nivel cientifico de suficiente relevancia y ha derivado en las siguientes publicaciones y comunicaciones:

Publicación:
E Rivin del Campo, S Rivera, M Martínez-Paredes et al. Assessment of the novel online delineation workshop dummy run approach using FALCON within a European multicentre trial in cervical cancer (RAIDs). Radiother Oncol. 124.1 (2017), pág. 130-138.

Oral communications

April 2013 III Congreso de jóvenes investigadores en formación de la Universidad de Córdoba. *Advanced Cervical Cancer: Analysis of Target Volume Delineation.* E. Rivin.

April 2013 10th Scientific and Medical Days of Institut Curie. *RAIDs in gynaecology.* M. Kamal and E. Rivin

April 2014 European Society for Radiotherapy & Oncology (ESTRO) 33 Congress. *Quality input of an online delineation workshop in advanced stage cervical cancer. Initial results.* E. Rivin del Campo, S. Rivera, M. Martínez-Paredes, et al.

Presentation at working group

April 2013 Rational molecular Asessments and Innovative Drug selection in advanced stage cervical cancer (RAIDs) 6 month steering committee within the International. Charité-Mayo Conference. (Institut Curie-all RAIDs partners). *Achieving standard chemoradiotherapy quality control.* E. Rivin, C. Haie-Meder, E. Deutsch, S. Rivera.

Poster:

November 2016 V Congreso de jóvenes investigadores en formación de la Universidad de Córdoba. *Radiotherapy in advanced cervical cancer: initial results of a delineation workshop of organs at risk and target volumes within a European Multicentre Trial.* E.Rivin del Campo, Fabiola Romero Ruperto, Amalia Palacios Eito, Eric Deutsch, María Martínez Paredes.

Por todo ello, se autoriza la presentación de la tesis doctoral.

Córdoba, _12_ de ___Diciembre__ de ___2017____

Firma del/de los director/es

Fdo.:____MARIA MARTÍNEZ PAREDES Fdo.: ____AMALIA PALACIOS EITO

Fdo.: _____ERIC DEUTSCH

Acknowledgements

To my colleagues who have inspired me.

Ever since I was a young girl I wanted to do medical research. Little did I know that I would wind up in this whirlwind, passionate field called Oncology, where only the sky is the limit to research possibilities!

I began to walk this path hand in hand with my mentor and thesis director Amalia Palacios Eito and a colleague working on her doctoral thesis at the time, Sonia García Cabezas. With them I learned that funding and supportive infrastructure isn't always fundamental (although very helpful) in research. Sometimes just good, sound research proposals addressing pertinent questions and a lot of hard work can prevail! I thank them for getting me involved very early on in research projects, and showing me the ropes.

I continued this path with the guidance of Pablo Pérez Martinez, with whom I learned to conceive, design, execute, analyse and communicate the results of a project, from A to Z, within a Masters program. It was key to understand all steps in the process of clinical research, and even nowadays this experience allows me to relate to ongoing research projects that I sometimes enter in different periods of their evolution. That experience has enabled me to identify which steps have been taken, to ease my way into the project providing what is needed at that time. Very often I find myself recalling important remarks he made about my project, which I apply to this day. I thank him for teaching me method.

However, I have not only been inspired by more experienced and well seasoned researchers. My fellow Master students, Gracia Quintana Navarro and Patricia Peña Orihuela were examples of passion and perseverance, while overcoming adversities encountered in their trials. They taught me that if things don't work out as planned, you don't give up, you just learn from the experience and find a new approach from another angle. I thank them for their support, and for teaching me to hang in there.

Although I had been hoping to develop my PhD thesis on a European level, never in my wildest dreams would I have expected to be offered such a wonderful opportunity. I thank Sofia Rivera for having spotted me, somehow recognised my potential after a brief 5 minute conversation, and Eric Deutsch for entrusting me with the radiotherapy quality control for the European FP7 project, which culminates in this thesis. I thank them both for their guidance, availability and support. They have accompanied me and mentored me every step of the way.

The technical part of this online delineation workshop (ODW) was supported by Miika

Palmu, whom I also thank for continued support during the analysis portion of this thesis. The medical docent aspect of the workshop was provided by Christine Haie-Meder and the late Renaud Mazeron, may he rest in peace. I thank them for teaching me many practical subtleties in cervical cancer brachytherapy, and how to effectively evaluate pertinent questions in brachytherapy and communicate the results.

I thank Suzy Scholl and Maud Kamal and the rest of the Rational molecular Assessments and Innovative Drug Selection (RAIDs) consortium, as well as the participants of the ODW, without whom this thesis would not be possible. I thank Andrea Slocker for her participation in the execution of this project.

I thank the statisticians involved in this project, Isabelle Borget and Philippe Hupé. The statistics were not at all straightforward, and several different approaches were used until the right one was identified. I especially thank Philippe Hupé for his excellent understanding of the endpoints, and for testing model after model until finding the one which best suited this trial, by taking into consideration all effects.

I thank the University of Cordoba for accepting my research project in this International PhD program. I thank it for an excellent experience, with solid support from the faculty as well as an outstanding organisation.

Lastly, I thank the directors of my thesis, Maria Martínez Paredes, Amalia Palacios Eito and Eric Deutsch for their incredible support from the scientific, moral and logistical point of view, being both efficient and dependable.

A mi madre, que siempre me ha alentado a avanzar. Siempre me ha inculcado que soy capaz de lo que me proponga.

A mi familia por su apoyo incondicional, por aportarme alegría y fuerza.

A mis amigos que siempre han estado ahí y a los nuevos que he encontrado durante esta aventura.

En especial a mi marido, que ha estado a mi lado durante este largo camino.

A mi hijo, parido prácticamente en paralelo a esta tesis. Por su paciencia durante mis horas delante de ella, y por su sonrisa.

A Europa. Porque la colaboración entre países es posible, es enriquecedora y no sólo es el presente, sino también el futuro.

This project has received funding from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement No 304810.

Acting the recipient for the grant for DUERTECC/EURONCO (Diplôme Universitaire Européen de Recherche Translationnelle Et Clinique en Cancérologie).

Contents

List of Figures

List of Tables

List of Acronyms

AAPM	American Association of Physicists in Medicine
BCG	Bacillus Calmette-Guérin
BQT	braquiterapia
BT	brachytherapy
CALGB	Radiation Oncology Committee for the original Cancer and Leukemia Group B
CBCT	cone-beam CT
CCLA	cáncer de cérvix localmente avanzado
CI	confidence interval
CIN	cervical intraepithelial neoplasia
C1	baseline contouring
CT	computed tomography
C3	final contouring
CTV	clinical target volume
CTV-node	nodal elective volume
CTV-p	GTV-p, uterus and vagina, at least 20 mm below GTV-p
C2	guideline contouring
DVH	dose volume histogram
EBRT	external beam radiotherapy
eCRFs	electronic case report forms

EMBRACE	An intErnational study on MRI guided BRachytherapy in locally Advaced CErvical cancer
EORTC	European Organisation for Research and Treatment of Cancer
ESTRO	European Society for Radiotherapy & Oncology
FALCON	Fellowship in Anatomical deLineation and CONtouring
FDR	false discovery rate
FIGO	International Federation of Gynecology and Obstetrics
Gyn GEC-ESTRO	Groupe Européen de Curiethérapie – European Society for Radiotherapy and Oncology
GHG	Global Clinical Trials RTQA Harmonization Group
GTV	gross tumor volume
GTV-node	radiologically pathological lymph nodes to boost
GTV-p	cervix, parametria and vaginal gross disease
Gy	Gray
HDR	high dose rate
HIV	human immunodeficiency virus
HPV	human papillomavirus
HR	hazard ratio
HR-CTV	high risk CTV
ICC	intra-class correlation
ICR	individual case review

ICRU	International Commission for Radiation Units and measurements.
IGBT	image guided brachytherapy
IGRT	image guided radiotherapy
IMPT	intensity modulated proton radiotherapy
IMRT	intensity modulated radiotherapy
INTERLACE	Induction Chemotherapy Plus Chemoradiation as First Line Treatment for Locally Advanced Cervical Cancer
IR-CTV	intermediate risk CTV
IROC	Imaging and Radiation Oncology Core Group
LACC	locally advanced cervical cancer
LDR	low dose rate
LVI	lymphovascular invasion
MRI	magnetic resonance imaging
NCI	National Cancer Institute
NCTN	National Clinical Trials Network
NNT	number needed to treat
OAR	organs at risk
ODW	online delineation workshop
Pap smear	papanicolau test
PDR	pulsed dose rate
PTV	planning treatment volume
QARC	Quality Assurance Review Center
RAIDs	Rational molecular Assessments and Innovative Drug Selection

ROG	Radiation Oncology Group
ROI	region of interest
RPC	Radiological Physics Center
RPPA	reverse phase protein arrays
RT	radiotherapy
RTE	radioterapia externa
RTOG	Radiation Therapy Oncology Group
RTQA	Radiation Therapy Quality Assurance
SBRT	stereotactic body radiation therapy
SEER	Surveillance, Epidemiology, and End Results
TDE	taller de delineación en línea
3D	three-dimensional
TRAK	total reference air kerma
TV	target volume
2D	two-dimensional
US	ultrasound
VODCA	Visualization and Organization of Data for Cancer Analysis

Notations

dice1 DICE value for baseline contouring (C1)

dice2 DICE value for guideline contouring (C2)

dice3 DICE value for final contouring (C3)

ROI.type type of ROI that has been contoured (TV = Target Volume, OAR = Organ At Risk)

part participant

inst institution from which the participants are from

exp experience of the participant (AI, AE, JR or SR)

pos position of the participant (A or R)

img imaging technique used by the institutions

slice slice number in the contour

slice2 square value of the slice number in the contour

C Contouring periods (baseline, guideline and final contouring)

ant anterior

post posterior

r right

l left

alatr anterolateral right

alatl anterolateral left

platr posterolateral right

platl posterolateral left

AE Experienced attending ($>$ 5 years of experience)

AI Less experienced attending ($\leq$ 5 years of experience)

SR Senior resident ($>$ 2 years of experience)

JR Junior resident ($\leq$ 2 years of experience)

, **, ** *p-value* $<$ 0.05, *p-value* $<$ 0.01, *p-value* $<$ 0.001 respectively

Resumen

Introducción - Las herramientas educativas en línea permiten una formación médica interactiva para estudiantes dispersos geográficamente. Pocos estudios tienen como objetivo primario su validación objetiva. Por sus ventajas logísticas, se utilizó un taller de delineación en línea (TDE) para homogeneizar el contorneo como control de calidad de radioterapia en el ensayo prospectivo multicéntrico europeo RAIDs (Rational molecular Assessments and Innovative Drug Selection) en cáncer de cérvix localmente avanzado (CCLA). Se realizaron dos TDE idénticos para evaluar sobre un paciente de simulacro las delineaciones de los centros RAIDs.

Objetivos

- Principales:

    1. Realizar una validación interobservadora de los TDE en CCLA mediante una evaluación cualitativa y cuantitativa de la mejoría de delineaciones entre participantes (variabilidad interobservadora).
    2. Efectuar una validación intraobservadora del TDE en CCLA mediante una evaluación cualitativa y cuantitativa de la mejoría de delineaciones de cada participante (variabilidad intraobservadora).

- Secundarios:

    1. Evaluar la metodología de enseñanza del TDE mediante cuestionarios de satisfacción.
    2. Análizar las delineaciones de los participantes de los centros europeos que utilizan la resonancia magnética para planificar braquiterapia (BQT) con respecto a los que no la usan.

Metodología - Se realizaron dos TDE incluyendo 46 especialistas de 14 centros de RAIDs. Se estableció una colaboración técnica con la Sociedad Europea de Oncología Radioterápica (ESTRO). La formación se realizó por una experta en la materia, CHM. Mediante 3 presentaciones en directo en línea se presentaron la plataforma de contorneo Fellowship

in Anatomical deLineation and CONtouring (FALCON) EduCase^{TM}, las guías de delineación y los contornos de los participantes. Los participantes completaron contornos basales (C1), guía (C2) y finales (C3) para radioterapia externa (RTE) y braquiterapia (BQT) en CCLA. La variabilidad interobservadora e intraobservadora se evaluó cuantitativamente (Índice DICE) y se analizaron mediante un modelo lineal mixto. La variabilidad intraobservadora fue evaluada cualitativamente mediante el test de McNemar.

Resultados - Nueve participantes enviaron contornos para RTE y BQT (C1-C3). Treinta y dos envió algún contorno. La comparación interobservadora cuantitativa de RTE mostró una mejoría significativa entre C2 y C1 para intestino, CTV ganglionar, CTV-p y GTV-ganglio con un detrimento significativo para GTV-ganglio (entre C3 y C1; C2), CTV-p (entre C3 y C2) e intestino (entre C3 y C2), es decir, una mejoría en general entre C2 y C1, con un detrimento significativo entre C3 y C2 en dos volúmenes blanco y un órgano de riesgo. Para BQT hubo una mejoría significativa entre C2 y C1 para vejiga, GTV, HR-CTV e IR-CTV con un detrimento significativo para vejiga (entre C3 y C2), en resumen, una mejoría general entre C2 y C1, con sólo un detrimento entre C3 y C2 para la vejiga. Las comparaciones intraobservadoras cuantitativas mostraron una mejoría significativa de delineaciones de regiones de interés entre C2 y C1, C3 y C1 y C3 y C2 para RTE y entre C2 y C1 para BQT. Las comparaciones cualitativas intraobservadoras destacables en BQT fueron una mejoría significativa en las direcciones derecha y posterolateral derecha para los volúmenes blanco entre C2 y C1, que pasaron a ser significativamente peores para estas direcciones entre C3 y C1; C2. El resultado de las preguntas acerca de la organización y el contenido del TDE del cuestionario de satisfacción (escala del 1 al 5, 1=poco satisfecho, 5=muy satisfecho) para los 20 participantes que respondieron fue de media 4.36 (rango: 3.95-4.60). Los centros que empleaban resonancia magnética para la delineación en BQT tuvieron una mejoría significativa para HR-CTV con respecto a los que usaban otras técnicas (entre C2 y C1: $p\text{-}value < 0.005$; entre C3 y C1: $p\text{-}value = 0.02$).

Conclusiones - Los TDE permiten formar, armonizar inicialmente la delineación y evaluar la experiencia de los centros antes de la inclusión de pacientes de una forma cómoda y eficaz en el seno de un estudio multicéntrico.

Summary

Introduction - E-learning programmes allow effective medical training for geographically dispersed participants. Few studies aim primarily to validate these programmes objectively. Considering its logistical aptitudes, an ODW using FALCON was used to homogenise delineation practice in cervical cancer as quality control among European centres participating in a multicentre prospective trial on locally advanced cervical cancer (LACC). Two identical ODW were performed, to evaluate the delineations of RAIDs centres on a fictitious patient.

Aims

- Primary endpoints:

    1. Interobserver validation of the ODW in LACC by using qualitative and quantitative assessments of the improvement of clinician contours between clinicians (interobserver variability).
    2. Intraobserver validation of the ODW in LACC by assessing qualitatively and quantitatively the improvement of the clinician contours for each clinician (intraobserver variability).

- Secondary endpoints:

    1. Evaluation of the teaching methodology of the ODW as reflected by the participant satisfaction questionnaires.
    2. Analysis of clinician contouring on magnetic resonance imaging (MRI) for European centres which use MRI for brachytherapy planning versus those which do not.

Methods - Two ODW included 46 clinicians from 14 RAIDs centres. A technical collaboration was established with European Society for Radiotherapy & Oncology (ESTRO). Training was performed by an expert in the field, CHM. Through 3 live online presentations,

the contouring platform FALCON EduCase^{TM}, the delineation guidelines and the participant contours were presented. Participants were asked to complete baseline contouring (C1), guideline contouring (C2) and final contouring (C3) for external beam radiotherapy (EBRT) and brachytherapy (BT) for LACC. Interobserver and intraobserver variability was quantitatively evaluated (DICE index) and analysed by a linear mixed model. Intraobserver variability was qualitatively evaluated by the McNemar test.

Results - Nine participants contoured for EBRT and BT for C1 - C3. Thirty-two clinicians submitted any contour.

Concerning interobserver quantitative comparisons for EBRT, significant improvement was observed for C2 vs. C1 for bowel, nodal elective volume (CTV-node), GTV-p, uterus and vagina, at least 20 mm below GTV-p (CTV-p) and radiologically pathological lymph nodes to boost (GTV-node), versus a significant detriment for GTV-node (C3 vs. C1; C2), CTV-p (C3 vs. C2) and bowel (C3 vs. C2), showing overall an improvement in C2 vs. C1, versus a detriment in C3 vs. C2 for two target volumes and one organ at risk. In the BT treatment there was significant improvement for C2 vs. C1 for bladder, gross tumor volume (GTV), high risk CTV (HR-CTV) and intermediate risk CTV (IR-CTV), versus a significant detriment for bladder (C3 vs. C2), thus a general improvement in C2 vs. C1, with only a detriment in C3 vs. C2 for bladder. As for intraobserver quantitative comparisons, a significant improvement was observed for contouring a region of interest between C2 vs. C1, C3 vs. C1 and C3 vs. C2 for EBRT and between C2 vs. C1 for BT. Notable intraobserver qualitative comparisons were found for BT, a significant improvement towards the right and posterolateral right directions for target volumes in C2 vs. C1, which became significantly worse in these directions in C3 vs. C1 and C2. The average result of the Organisation and Content items in the satisfaction questionnaire for the 20 ODW participants who responded (scale 1-5, 1=poor, 5=excellent) was 4.36 (range 3.95-4.60). With regard to the imaging technique used for BT planning, centres using magnetic resonance imaging (MRI) did significantly better in the BT case for HR-CTV than centres using other techniques (C2 vs. C1: $p\text{-}value < 0.005$; C3 vs. C1: $p\text{-}value = 0.02$).

Conclusions - ODW allow to train, initially harmonise contouring and assess the experience of centres before patient inclusion conveniently and efficiently in the context of a multicentre trial.

1 Introduction

Contents

1.1 General context: cervical cancer

1.1.1 Epidemiology

Even nowadays, when screening programmes for cervical cancer are readily accessible to women in developed countries, this illness still accounts for a non-negligible morbidity and mortality in Europe. Ferlay et al., (2015) have found it ranks globally as the fourth in Europe and as the sixth most frequent cancer in women. When considering emerging countries, it is currently the second most common female cancer, as it used to be worldwide in 1975. There were an estimated 58,400 new cases in 2012 in Europe, with 33,900 occurring in Central and Eastern Europe. In terms of mortality, the estimated number of deaths from cervical cancer in 2012 was 24,400 women in Europe, of which 15,400 were in Central and Eastern Europe, and only 3,500 in Western Europe. This data highlights a notable imbalance between these European regions.

The EUROCARE-5 study investigated the imbalance in survival among European regions for female breast and genital cancers. For cervical cancer, they found poor survival in Eastern Europe with an augmentation of the incidence in several Eastern European countries, as opposed to an incidence which was either reduced or maintained stable in the majority of other European countries. The authors believe these differences might be due to either

deficits in screening/prompt diagnosis, suboptimal treatment, or a combination of both (Sant et al., 2015).

A recent publication which considered inferences for Eastern European countries based on the epidemiological changes of cervical cancer over time in Bulgaria suggested that other Eastern European countries adopt the actions taken there to decline its incidence, such as the intensification of screening (Samson et al., 2016). They even consider a common prevention programme for all Eastern European countries. Another of their proposals was for Eastern European countries to improve treatment accessibility in order to achieve lower rates of mortality.

Survival has been found to be clearly linked with the International Federation of Gynecology and Obstetrics (FIGO) classification staging at diagnosis, with rates varying between 93% survival in stage IA and 15% in stage IVb (Reuzé et al., 2017).

Several risk factors have been linked to cervical cancer, such as oral contraception (used over 5 years), having a transplanted organ, human immunodeficiency virus (HIV) infection, active or passive smoking, engaging in sexual intercourse at an early age, having a large amount of sexual partners, multiparity and a family member of the first degree affected by cervical cancer. These factors may favor infection by the human papillomavirus (HPV), which causes the vast majority of cervical cancer cases, or facilitate the development of cervical cancer (Hillemanns et al., 2016; Lea and Lin, 2012). In the presence of a sustained infection by HPV, the average time between viral acquisition and the appearance of cervical intraepithelial neoplasia (CIN) and ultimately cervical cancer is of 15 years (Lea and Lin, 2012). The development of cervical cancer can also be due to female hormones, with most cases presenting above 45 years of age (Plummer et al., 2012; Pike et al., 2004).

HPV vaccine

The first widely used HPV vaccine only included subtypes 16 and 18 (some also 6 and 11). Although approximately 70% of cervical cancer cases are caused by subtypes 16 and 18, there was still a need to increase the coverage to prevent even more cases (Hillemanns et al., 2016). Thus, in 2015, the European Medicines Agency recommendations included a new vaccine ensuring protection from 9 HPV subtypes causing 90% of cervical cancer cases (Joura et al., 2015). The other advantage of this vaccine is that it is administered in 2 as opposed to 3 doses, thus facilitating adherence. Currently, the rates of women vaccinated differ greatly between European countries. The amount of national school vaccination plans oscillate

between under 30% and over 80% (Hillemanns et al., 2016). These school programmes are of utmost importance as vaccination has better results before sexarche (indication in women between 9 and 26 years old). In spite of HPV vaccination, these women still must be screened routinely by the papanicolau test (Pap smear), as unfortunately not all HPV subtypes which may induce cervical cancer are covered (Lea and Lin, 2012).

Screening

Although prevention by vaccination is beginning to play a role in reducing the incidence of cervical cancer, screening programmes are still quite necessary. Traditional routine screening has consisted of the Pap smear technique, which does not have a high sensitivity (55% to 80%) and only detects accurately stage I cervical cancer in 30-50% of cases (Benoit et al., 1984). To improve this technique an option is liquid-based cytology, allowing better sample quality, therefore improving its interpretation in less time (Hillemanns et al., 2016).

Recently, HPV testing has been introduced in some national screening programmes, for instance, Germany. This is based on the evidence drawn from long term follow up of patients included in four major randomised European trials initially powered to distinguish differences in detection of precancerous cervical lesions between the Pap smear and HPV testing techniques (Ronco et al., 2014). With a median follow up of 6.5 years, the pooled results (intention to screen) of the Dutch (POBASCAM), Italian (NTCC), Swedish (Swedescreen) and British (ARTISTIC) studies were in favour of HPV testing for screening, protecting from cervical cancer 60-70% better than the cytology technique. This is due to a higher diagnosis of precancerous lesions at first screening, before progression. As new HPV infections often are transitory and CIN lesions in regression may be overdiagnosed, the time between HPV-screenings may be incremented to 5 years, resulting in less colposcopies and biopsies than with cytologies (more cost-effective) (Ronco et al., 2014). An interesting point was raised in the ARTISTIC trial (Kitchener et al., 2008)when assessing psychological distress amongst study subjects (Generalised Health Questionnaire-28). There was no significant difference in distress between patients who had an HPV-test and a cytology versus those who only had a cytology (Kitchener et al., 2008). Psychological distress is an important factor to consider in screening programmes, and these results are encouraging for the use of HPV-testing in cervical cancer screening from the psychosocial point of view.

1.1.2 Treatment modalities

Current guidelines include surgical or radiotherapy options for early stage cervical cancer, up to stage IB1 (Koh et al., 2017). Stage IA1 may be treated by conization or simple hysterectomy, but if there is lymphovascular invasion (LVI) there is a slightly larger risk of lymphatic spread (5%), thus a modified radical hysterectomy with pelvic lymphadenectomy is recommended (Lea and Lin, 2012; Koh et al., 2017). From two factors indicating higher risk of recurrence (LVI, large tumour size, deep invasion of stroma) an option is to administer adjuvant pelvic radiation therapy with or without concurrent cisplatin (Haie-Meder et al., 2010a). In stage IA2 cervical cancer standard treatment is modified radical hysterectomy with pelvic lymphadenectomy, although fertility sparing techniques include conization or trachelectomy. When unfortunately surgery results in margins which are positive, or evidences involvement of either the parametria or pelvic lymph nodes, there is an indication for postoperative radiochemotherapy in stage IA1 and IA2 tumours (Haie-Meder et al., 2010a). In both stages IA1 and IA2 without LVI, in unsuitable candidates for surgery, an alternative is brachytherapy (BT) (Lea and Lin, 2012; Koh et al., 2017). Several treatment options may be proposed to treat stage IB1 cervical cancer. The surgical option is radical hysterectomy with bilateral oophorectomy and pelvic lymph node dissection. Fertility sparing treatment by radical trachelectomy is also a possibility in young patients with very good prognostic features (smaller than 2 cm, no LVI and negative lymph nodes). Another option is pelvic external beam radiotherapy (EBRT) with or without concurrent chemotherapy followed by intracavitary BT (Haie-Meder et al., 2010a; Koh et al., 2017). A last possibility is either preoperative BT with surgery after 6-8 weeks or radiochemotherapy, BT and surgery (Haie-Meder et al., 2010a; Koh et al., 2017). As in stages IA1 and IA2, in those patients who were operated in first intent, but whose surgical margins were positive, or had parametrial or pelvic lymph node involvement, postoperative radiochemotherapy should be administered (Haie-Meder et al., 2010a).

Concurrent chemoradiotherapy is the standard of care for locally advanced cervical cancer (LACC), FIGO stages IB2 through IVA (Haie-Meder et al., 2010a; Koh et al., 2017). As this thesis was centred on LACC, this section will elaborate on its treatment. Five different randomised trials have found a reduction in the risk of death of 30-50% when concurrent chemoradiotherapy using cisplatin was administered (**Table 1.1**) (Koh et al., 2017). A meta-analysis of 18 studies including data from 3,452 patients showed that in the 13 trials which compared the same radiotherapy treatment with chemoradiation there was a 6% higher

5-year survival, with an even higher 5-year survival improvement of 19% in two studies which added adjuvant chemotherapy after chemoradiation (Vale, 2008). This indication was reinforced by the results from a registry of the Canadian population, including 4,069 patients diagnosed with cervical cancer. Three-year overall survival went from 58.6% in the time period when patients received radiation alone (1995-1998) to 69.8% when they received chemoradiation (1999-2001) (Pearcey et al., 2007). Thus, the survival benefit with chemoradiation has been thoroughly established, which outweighs its higher rates of acute toxicity, especially haematological and of the digestive tract (Haie-Meder et al., 2010a).

Study	n	Median follow-up	FIGO	Control	Comparison	RR for death
Stehman et al., (2007)	369	8.4	IB2	RT	RT + weekly cisplatin	0.63
Rose et al., (2007)	526	8.8	IIB-IVA	RT + Hydroxyurea	RT + weekly cisplatin RT + cisplatin + 5-FU + hydroxyurea	0.57 0.51
Eifel et al., (2004)	228	6.6	IB-IVA	Extended field RT	RT + cisplatin + 5-FU	0.49
Whitney et al., (1999)	368	8.7	IIB-IVA	RT + Hydroxyurea	RT + cisplatin + 5-FU	0.72
W. A. Peters et al., (2000)	243	3.5	IB or IIA (post-operative selection)	RT	RT + cisplatin + 5-FU	0.50

Table 1.1: Estimates of the Relative Risk (RR) of Death in Five Clinical Trials of Concurrent Chemotherapy and Radiotherapy (adapted from Koh et al., (2017)).

Chemoradiation is administered by EBRT followed by intracavitary BT, with concurrent chemotherapy, most often cisplatin-based. Improvements in each of these treatments are leading to the optimisation of this therapy. Thus, the use of intensity modulated radiotherapy (IMRT) reduces toxicity of the digestive tract while allowing dose escalation (Marnitz et al., 2011; Gandhi et al., 2013). The most relevant improvement in the BT field has been the incorporation of image guided brachytherapy (IGBT), which not only reduces the toxicity of BT, but also allows the administration of a higher dose, leading to better local control (Sturdza et al., 2016; Tanderup et al., 2016). Research also focuses on improving chemotherapy treatment, by both defining the best drug combination and treatment sequence (Haie-Meder et al., 2010a; Vordermark, 2016).

EBRT for cervical cancer has been administered by three-dimensional (3D) conformal radiotherapy in most centres since the turn of the 21st century. However, IMRT has gradually become the preferred technique to administer EBRT for these patients, based on the following clinical evidence. The fields included in this treatment are the cervical tumour, uterus, parametria, part of the vagina (the amount depends on vaginal tumour extension) and

pelvic lymph nodes +/- inguinal and or paraaortic lymph nodes when involved. The total dose administered is between 45 and 50.4 Gray (Gy) in fractions of 1.8-2 Gy (Vordermark, 2016; Gandhi et al., 2013). Gandhi et al., (2013) compared toxicity and disease-free and overall survival in a small randomised trial including 44 patients, between 3D conformal radiotherapy and IMRT, both with concomitant cisplatin, for LACC. Pelvic radiotherapy with or without extended fields, when indicated, was administered. Although there was no significant difference in terms of survival, there were notable differences in favour of the IMRT arm for both acute and late digestive tract toxicity. There were 32% less grade 2 and almost 23% less grade 3 acute gastrointestinal toxicities, and 36.4% less late gastrointestinal toxicities in the IMRT arm (Gandhi et al., 2013). Another advantage of using IMRT for the treatment of cervical cancer is the ability to administer a simultaneous integrated boost to macroscopic lesions. The feasibility of this technique was presented in a study comparing IMRT by helical tomotherapy with standard IMRT (Marnitz et al., 2011). The technique was feasible for both IMRT techniques, administering a simultaneous integrated boost at a dose of 2.12 Gy per fraction, 28 fractions, to the bilateral parametria.

BT is an essential treatment in LACC, as has been shown by an analysis of the Surveillance, Epidemiology, and End Results (SEER) database of patients treated for stage IB2-IVA carcinoma of the cervix (Han et al., 2013). The multivariate analysis of the cohort matched by propensity-score demonstrated that treatment with EBRT + BT independently induced higher cause specific survival and overall survival, with a hazard ratio (HR) of 0.65 (95% confidence interval (CI): 0.57-0.71) and 0.66 (95% CI: 0.60-0.74), respectively, when compared with EBRT alone. Interestingly, the year when Medicare insurance began covering IMRT in each SEER region, there was a clear decrease in the use of BT. While, as mentioned earlier, IMRT allows better sparing of organs at risk (OAR) than 3D conformal EBRT, it does not allow the administration of such high doses to the tumour as BT while avoiding OAR. This was shown in a study comparing magnetic resonance imaging (MRI) based IGBT with IMRT and intensity modulated proton radiotherapy (IMPT) (Georg et al., 2008). This study was particularly interesting, as it included IMPT, which has a steep dose gradient, allowing for better target coverage while sparing OAR. However, the study showed that IGBT was superior not only to IMRT, but also to IMPT, with better coverage of the gross tumor volume (GTV) (**Figure 1.1**). High dose rate BT treatments after EBRT for cervical cancer are most often administered in 3-5 fractions of 5-7 Gy per fraction (Vordermark, 2016).

Classic two-dimensional (2D) BT treatments were prescribed to *point A*, as designated by the *Manchester System* (Vordermark, 2016). The drawback of this system is that it did not

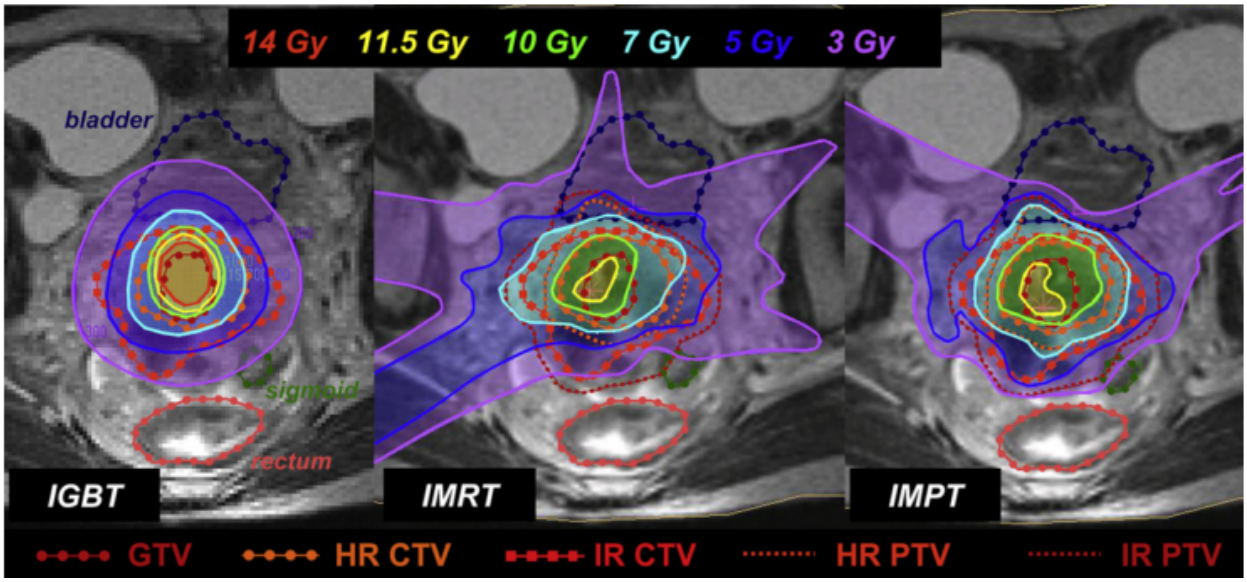


Figure 1.1: Isodose curves for IGBT, IMRT and IMPT. Abbreviations: HR-CTV: High risk CTV; IR-CTV: Intermediate risk CTV; HR-PTV: High risk planning treatment volume; IR-PTV: Intermediate risk planning treatment volume. Reprinted from D. Georg et al. **Image-guided radiotherapy for cervix cancer: high-tech external beam therapy versus high-tech brachytherapy**. *Int. J. Radiation Oncology Biol. Phys.* 71.4 (2008), pp. 1272–1278, © 2008, with permission from Elsevier.

take into account anatomic variations between patients nor the extension of the tumour. The need to determine more accurately the coverage of the tumour, as well as the dose received by the OAR, led to the development of 3D IGBT, either computed tomography (CT)-based, or ideally MRI-based. Much effort has been made by the Groupe Européen de Curiethérapie – European Society for Radiotherapy and Oncology (Gyn GEC-ESTRO) to optimise IGBT and develop guidelines for its implementation. The language to be used for IGBT was introduced in Gyn GEC-ESTRO guidelines published in 2005, where the terminology to be applied in this technique was clearly defined (Haie-Meder et al., 2005). Thus, GTV, high risk CTV (HR-CTV) and intermediate risk CTV (IR-CTV) delineation definitions were established. The aim of these guidelines was to improve contouring accuracy of target volumes in IGBT because it affects the treatment plan, as does contouring of OAR. For this, the importance of the use of MRI to improve target volume (TV) definition, especially for the GTV, was stressed. The second part of the Gyn GEC-ESTRO guidelines focused on dose volume histogram (DVH) values for TV as well as constraints for OAR (Pötter et al., 2006).

An intErnational study on MRI guided BRachytherapy in locally Advaced CErvical cancer (EMBRACE)[1] began in 2008 to prospectively evaluate observationally the clinical results of the use of MRI-based IGBT in several centres following Gyn GEC-ESTRO contouring and reporting guidelines. While we await the outcome of this trial, which closed to inclusions at

[1] http://www.embracestudy.dk

the end of 2015, the retroEMBRACE[2] data is available (Sturdza et al., 2016). RetroEMBRACE is the cohort of 731 patients for analysis which received IGBT in the monoinstitutional setting within 12 centres before they entered EMBRACE. With a median follow-up of 43 months, 5 year actuarial local control was 89% with a pelvic control of 84%, an overall survival of 65% and a cancer specific survival of 73%. The actuarial 5 year morbidity was 5% for bladder and vagina and 7% for the gastrointestinal tract (Sturdza et al., 2016). Though retrospective, and only 81% had MRI-based IGBT, these favourable results will hopefully be reconfirmed or even improved with the EMBRACE data.

An important aspect of standard treatment for LACC is keeping the overall treatment time between the beginning of chemoradiation and the end of BT under 55 days (Haie-Meder et al., 2010a). This aspect was evaluated in a subcohort of retroEMBRACE treated with MRI-based image guided radiotherapy (IGRT) (Tanderup et al., 2016). In these 488 patients not only prolonged overall treatment time ($p\text{-}value = 0.004$, HR = 1.023 per day), but also larger HR-CTV volume ($p\text{-}value = 0.004$, HR = 1.017 per cm^3) were found to associate with lower local control rates. This study recommended adding to the dose which covers 90% of the HR-CTV of 85 Gy a dose of 5 Gy to reach equivalent 3 year local control per additional week of overall treatment time (>49 days median overall treatment time) or per 10 cm^3 of increased HR-CTV volume (Tanderup et al., 2016).

Recently, data has begun to become available for the use of stereotactic body radiation therapy (SBRT) in gynaecological cancers. It is quite useful to treat node metastases, whether pelvic and/or para-aortic, as well as for salvage treatment of extra-nodal recurrences in the pelvis. SBRT's most prominent role in gynaecological cancer is nodal boosting, as reported in a recent systematic review (**Figure 1.2**) (Mendez et al., 2017; Choi et al., 2009). SBRT for nodal boosting in 183 patients from 6 studies yielded a combined local control rate of 83% (median follow up range: 4-20 months) with under 5% of grade 3-4 toxicity of the gastrointestinal tract and only one patient with grade 3 genitourinary toxicity. When used as salvage treatment for pelvic recurrences in 10 studies, 73 patients in total, local control reported for 57 of those patients was 86% with a range of median follow up of 4-20 months. A less employed SBRT approach in cervical cancer is either radical treatment or postoperative. Seven studies present data of SBRT as a radical treatment in LACC, reporting results of 34 patients. With a median follow up of 6-22 months, they found a combined local control of 91%. Only 2 patients suffered grade 3 or 4 gastrointestinal side effects (Mendez et al., 2017).

A final aspect to be considered when discussing radical radiochemotherapy treatment for

[2]http://www.retroembrace.com

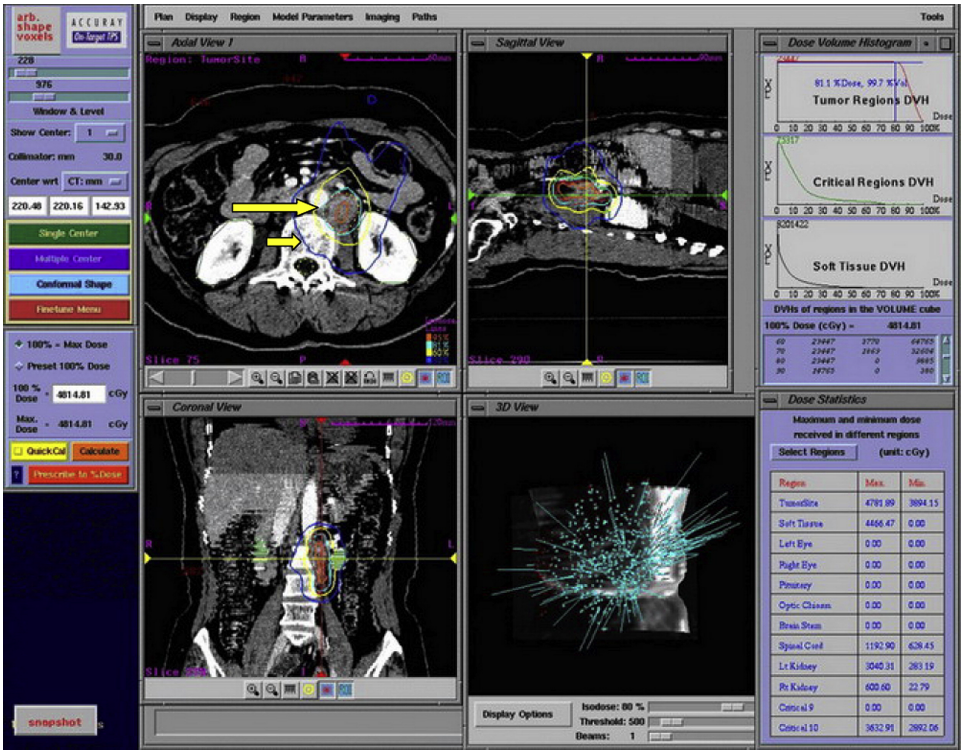


Figure 1.2: SBRT treatment plan for CyberKnife (39 Gy in three 13 Gy fractions). The GTV was the macroscopic para-aortic lymph node mass on CT scan (in red). The dose prescription was to the 81% isodose line of the maximum dose to cover the GTV + 2-mm margin (light blue, long arrow). The 30% isodose line is in blue (short arrow). Reprinted from C. Choi et al. **Image-guided stereotactic body radiation therapy in patients with isolated para-aortic lymph node metastases from uterine cervical and corpus cancer**. *International Journal of Radiation Oncology Biology Physics* 74 (2009), pp. 147–153 © 2009, with permission from Elsevier.

LACC is how to estimate the total dose (including posterior BT) to be received by surrounding OAR when boosting macroscopic pelvic lymph nodes with IMRT before BT. To this end, a recent study developed the concept of coverage probability (Ramlov et al., 2017). Coverage probability planning informs better of the spatial uncertainties in treatment administration, allowing for less stringent planning goals around the planning treatment volume (PTV), reducing the dose near OARs while still accurately boosting the macroscopic lymph nodes. The simultaneous integrated boost dose, to be administered to the macroscopically involved nodes during the administration of 45 Gy in 25 fractions to the pelvic lymph nodes, was 55-57.5 Gy depending on the prospective BT dose contribution. They reviewed 25 patients with 47 boosted lymph nodes. Boosted lymph nodes were contoured on cone-beam CT (CBCT), and the accumulated dose (EBRT + BT) they received was >57 Gy in 98% of the treated nodes (GTV nodes as contoured on the CBCT). When planning was performed per the coverage probability technique, the volume of the OAR (body, bowel and bones) receiving 50 Gy was significantly lower than when planning classically per PTV node coverage (Ramlov et al., 2017).

Considering the solid evidence in favour of concomitant chemoradiation, especially using cisplatin, as previously mentioned, improvements to this scheme are under investigation. Efforts in this direction have consisted of either an intensified chemotherapy dose approach to chemoradiation, complementary sequential chemotherapy, or both (Vordermark, 2016). Another approach, the use of targeted therapy, was contemplated as an ulterior application of the results obtained in the European multicentre study within which this thesis was conducted.

A study intensifying both chemoradiotherapy and adding sequential chemotherapy compared standard cisplatin-based chemoradiation to cisplatin-gemcitabine chemoradiation followed by two adjuvant cycles of cisplatin-gemcitabine (Dueñas-González et al., 2011). Even though the experimental arm showed higher 3 year progression free survival, by almost 10%, and higher overall survival, there was almost 3 times more grade 3/4 haematologic toxicity in the cisplatin-gemcitabine arm. A main issue in this trial is that it was run, for the most part, in developing countries. The radiotherapy treatments were not optimal, as neither IMRT (possibly even 2D EBRT was used for some patients, as field borders were specified in the supplement) nor IGBT were used. Furthermore, no centralised quality assurance was performed. This puts the results in question, thus this scheme is not widely used (Vordermark, 2016).

As for the use of targeted therapy in LACC, vascular endothelial growth factor targeting

with bevacizumab has been included in clinical trials within a chemoradiotherapy schedule. The RTOG 0417 trial used a cisplatin-bevacizumab combination concomitantly with standard EBRT, followed by BT (Schefter et al., 2012). This combination proved to be safe, and when looking at the 3-year overall survival in this trial, and the chemoradiation arm of RTOG 90-01, results when adding bevacizumab were 81.3% vs. 76.8% (Schefter et al., 2014).

Another direction under study is neoadjuvant chemotherapy in LACC. A systematic review performed in 2003 including a meta-analysis of individual patient data found a very significant decrease in the risk of death with neoadjuvant chemotherapy before surgery compared with radiotherapy alone (Neoadjuvant Chemotherapy for Cervical Cancer Meta-analysis Collaboration, 2003). However, it was not compared with the current standard of care, which is chemoradiation, and the GOG 141 was negative, thus no improvement was shown with vincristine-cisplatin neoadjuvant chemotherapy followed by surgery in bulky stage IB cervical cancer. Hence, neoadjuvant chemotherapy has not been broadly adopted in LACC (Haie-Meder et al., 2010a). Results from the 2003 meta-analysis did provide rationale for the phase III EORTC 55994 trial which randomises patients with stage IB or II cervical cancer to either neoadjuvant chemotherapy followed by surgery or chemoradiotherapy. The results are awaited for 2019. A final remark with respect to this subject is based on a recent review from a Brazilian team (de Azevedo et al., 2016). Considering that LACC is much more prevalent in developing countries, many of which have limited access to radiation therapy, they decided to revise the current evidence available concerning neoadjuvant chemotherapy in this setting. Their review of 7 pertinent trials concluded that upfront neoadjuvant chemotherapy may be an option for those patients which have a lower risk of progression, especially in the context of countries with less radiotherapy availability (de Azevedo et al., 2016).

1.2 Specific context (RAIDs)

This thesis was performed within the frame of the European multicentre study Rational molecular Assessments and Innovative Drug Selection (RAIDs) in Advanced Stage Cervical Cancer, funded by the European Commission (Seventh Framework Programme) (Consortium RAIDs, 2017). RAIDs integrated genomic studies and biochemical, molecular, virological and cellular biology investigations on cervical cancer cells and tissues. Academic clinical centres, small-medium enterprises, and academic and translational research platforms par-

ticipated in this project (**Figure 1.3**). The 22 clinical centres were from Western and Eastern European countries (**Table 1.2**). As stated earlier, there is a large disparity between certain European regions in the incidence and mortality of cervical cancer, due to differences in screening programme intensity and access to treatment, as well as the adequacy of the treatment. Thus, one of RAIDs goals was to homogenise clinical management of cervical cancer across the participating European centres.

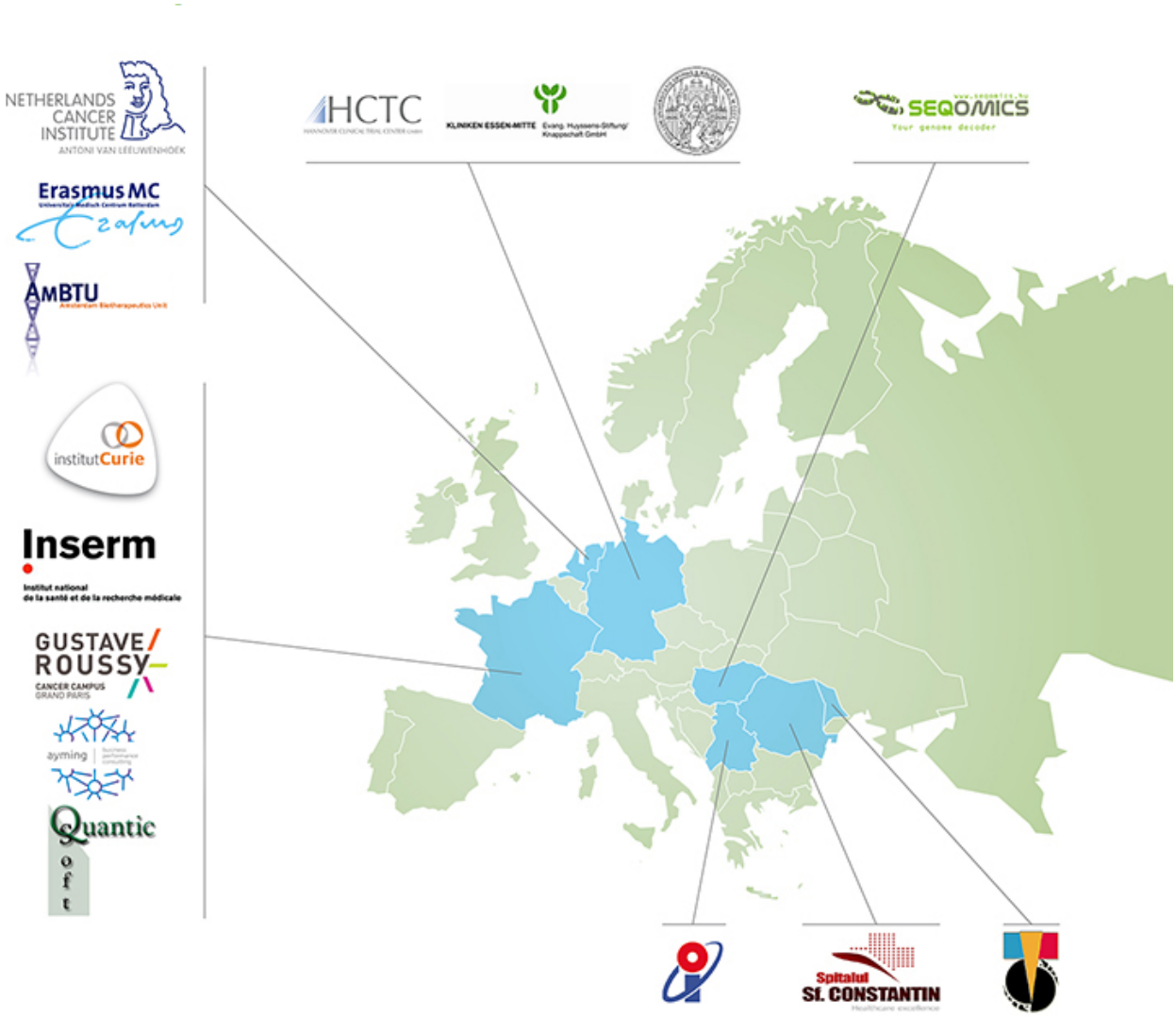


Figure 1.3: European map showing participating RAIDs consortium centres.

The main objective of RAIDs was to define stratification criteria for therapy by determining the tumour's molecular profile (exome sequencing, reverse phase protein arrays (RPPA), and integrative bioinformatics analyses). Tumour material was assessed within clinical and translational trials before treatment and following therapeutic HPV vaccination.

Before performing these clinical and translational trials, treatment homogeneity needed to be achieved among centres, particularly in radiochemotherapy, standard treatment for

LACC. This was the aim of the task: *Quality control for standard therapy*, in particular radiotherapy (RT) management in all clinical centres for which Gustave Roussy Cancer Campus was the leader. This thesis was developed in partnership with this centre.

Considering the geographical difficulties involved and the limited funding, the best option for training of the different centres in radiotherapy treatment contouring for LACC was an online workshop. This also allowed Gustave Roussy Cancer Campus to contribute to another task within RAIDs: *Training for young doctors and specialised health care physicians as well as other health care actors.*

FRANCE	Institut Bergonié, Bordeaux
	Institut Curie, Saint Cloud
	Institut Curie, Paris
	Institut Gustave Roussy, Villejuif
	Institut de Cancérologie de Lorraine, Nancy
	Centre Paul Strauss, Strasbourg
	Centre Georges François Leclerc, Dijon
	Institut de Cancérologie de l'Ouest, Nantes
	Institut de Cancérologie de l'Ouest Paul Papin, Angers
	Hôpital Européen Georges Pompidou, Paris
	Institut du Cancer de Montpellier, Montpellier
	Centre Antoine Lacassagne, Nice
SERBIA	Institute of oncology of Vojvodina
NETHERLANDS	Netherlands Cancer Institute – Antoni van Leeuwenhoek (NKI-AVL), Amsterdam
	Amsterdam Medical Center (AMC), Amsterdam
GERMANY	Hannover Medical School(MHH) Universitätsklinikum C.G. Carus_Klinik & Poliklinik für Frauenheilkunde & Geburtshilfe, Dresden
ROMANIA	Teo Health S.A., Brasov
	Spitalul Clinic Municipal, Oradea
	Institutul Regional de Oncologie, Lasi
	Clinica Radioterapie, Timisoara
MOLDOVA	Moldavian Oncological Institute, Chisinau

Table 1.2: List of participating RAIDs clinical centres.

1.3 Quality control in radiotherapy

Researchers began to realise the need for quality control for radiotherapy within clinical trials at the end of the 1970's, with an initial aim to increase the number of patients included in trials which were evaluable per protocol (Fairchild et al., 2013). At this time, the Radiation Therapy Oncology Group (RTOG) implemented sample case reviews for radiotherapy treatments (FitzGerald, 2012). Thus, in 1982, outcome results with respect to Radiation Therapy Quality Assurance (RTQA) for Southwest Oncology Group Protocol 7628, ran-

domising small cell lung cancer patients between multiagent chemotherapy and radiation therapy with or without Bacillus Calmette-Guérin (BCG), were presented. They showed that patient cases with major variations from protocol had significantly worse survival (40 vs. 60 weeks, respectively; $p\text{-}value = 0.002$), a lower improvement of the response rate post-induction chemotherapy (27 versus 48%, respectively; $p\text{-}value = 0.05$) and a higher rate of chest recurrences (77 vs. 55%; $p\text{-}value = 0.047$) than protocol compliant patients (White et al., 1982). These early results for the impact of quality assurance support the affirmation by Poortmans et al., (2005): the fact that patients included in clinical trials tend to have better outcomes than others with similar characteristics, both in terms of therapeutic results and less complications, with a better quality of life, can be due to patient selection and more precise treatment, thanks to diagnostic and therapeutic quality assurance and more attentiveness by the health care staff in centres involved in clinical trials. Though radiotherapy treatments which are not per protocol may waste time, effort and money and even harm patients, evidence shows that participating in RTQA programmes ameliorates not only treatment delivery for trial patients, but also for off-trial patients (Weber et al., 2011; Haworth et al., 2009).

The first centre dedicated to quality assurance programmes was developed in the United States: the Radiological Physics Center (RPC). It has received funds from the National Cancer Institute (NCI) since the late 1960's to audit the dosimetries of participating centres of NCI cooperative clinical trials, and as aforementioned, a decade later the RTOG created radiotherapy sample case reviews. Also, the Quality Assurance Review Center (QARC) was created, within the Radiation Oncology Committee for the original Cancer and Leukemia Group B (CALGB) in 1976. The QARC elaborated a process to collect radiotherapy data for review, allowing study investigators to perform both interventional and retrospective reviews of radiotherapy treatments for all patients involved in clinical trials including this radiotherapy review data. Early results from the CALGB RTQA procedures showed treatment approaches from different centres were inhomogeneous, manifesting a need for monitoring of these treatments, especially when participating centres used techniques within the clinical trial which were different from their standard techniques (Fairchild et al., 2013). By the 1980's RTQA faced another challenge, due to the gradual implementation of 3D treatment planning, thanks to a more generalised access to CT. This led to the introduction of benchmarking exercises, allowing assessment of the radiotherapy equipment, the team and their capabilities. These exercises gave RTQA facilities insight as to how patients were treated in each participating centre (FitzGerald, 2012).

In Europe the European Organisation for Research and Treatment of Cancer (EORTC) RTQA programme was inaugurated in 1982 (Kouloulias, 2003). It was developed within the Radiation Oncology Group (ROG), which established the ways and the principles of RTQA within European clinical trials (Fairchild et al., 2012b).

To elaborate on the concept of RTQA, it allows for close observation and assessment of how radiotherapy techniques are performed within clinical trials, but also in an institution when used for internal quality assurance. Radiotherapy treatment preparation involves several professionals other than the treating physician, such as the radiation therapy technologist who scans the patient (in 3D EBRT), medical residents, physicists and lastly the radiation therapy technologist who administers the treatment. Consequently, human error and/or software faults may ensue during any of the steps of the process. RTQA has been defined as the procedures which guarantee consistency of the radiotherapy prescription and the safe application of the prescription through the treatment plan by administering the correct dose to the TV while sparing at best the OAR, minimising the radiation exposure of the personnel and providing adequate patient follow up during treatment as well as afterwards, to assess outcomes (Institute of Radiation Hygiene and World Heath Organization, 1988). RTQA has three pillars, the *structure* in which the treatment is delivered, the *process* of treatment preparation and the *outcome* for the patients (van der Schueren et al., 1993). The goals of RTQA are to define acceptable protocol deviations, identify possible reasons for larger deviations, and devising mechanisms to correct and prevent them, so as to decrease variability and uncertainties encountered during the steps of treatment planning and administration (Weber et al., 2011).

Technological innovations inducing advanced radiotherapy techniques have brought about a growing interest in RTQA programmes. The application of novel, more complex, radiotherapy treatment techniques increase the possibility of issues with the quality of the treatment delivered (Spry et al., 2008). Due to this, the cost of conducting clinical trials and the higher number of patients accrued, a larger demand has arisen for rigorous RTQA programmes which may ensure patients will be treated optimally, per protocol. In current times, only studies which are well conducted with a well documented RTQA programme have credibility. This is the only way to produce robust, definitive results which may be generalised and result in a change of clinical practice. Also of utmost importance is the fact that many modern trials are multidisciplinary and international, thus intrinsically less concordant and homogeneous across participating centres (Weber et al., 2011).

RTQA programmes within multicentre trials should:

- Clarify any protocol ambiguities with possible impact on treatment delivery.
- Train participating centres in radiotherapy aspects of the trial guidelines.
- Ensure treatment homogenisation between centres.
- Warrant sites satisfy the needed technical and staff requirements.
- Guarantee data integrity as well as its accuracy.
- Estimate variability between patients and institutions.
- Identify and correct any flaws in the design of the study (Weber et al., 2011).

RTQA programmes have come a long way since their initial conception in the 1960's-70's. They have not only kept up with cutting edge technological advances in the field of radiotherapy (and helped ease their implementation safely in departments within clinical trials), but have also integrated these and other advances in the RTQA process itself, allowing for a more streamlined, user-friendly RTQA programme. Initial efforts consisted of site visits, assessments of the personnel and infrastructure of participating centres, mechanical and dosimetric evaluations of treatment units, patient chart and portal imaging reviews, dosimetry audits by mailed in thermoluminescent dosimeters and even the development of radiobiological models to assess inter-institutional differences (Poortmans et al., 2005; Fairchild et al., 2012b). These assessments brought about the creation of the *dummy run*. This initially consisted of treating a dummy patient (an anatomical phantom) per protocol guidelines and was first used in the EORTC RTQA programme in the EORTC 22791 protocol, studying conventional vs. bifractional radiotherapy in cancers of the oropharynx. They found an accurate delivered dose in the central axis in all cases, but many differences in treatment planning with approximately one third of the cases with inadequate TVs (either too small or too large). Thus, the EORTC RTQA programme implemented this *dummy run* to be performed before opening trial centres to detect any possible deviations and correct any misinterpretation of the trial guidelines prior to patient inclusion. A couple years later, they decided to include case reviews of patients treated on the trial, individual case review (ICR), to detect random errors. These ICR consisted of a review of parameters of the patient, the tumour and the treatment to assess whether the patient had been treated per protocol guidelines (Poortmans et al., 2005). Several reasons which may be responsible for non-compliance with *dummy run* guidelines have been identified (**Table 1.3**) (Fairchild et al., 2012a).

Reason
Misinterpretation of or ambiguities in protocol instructions
Institutional differences in available resources
Different interpretations of clinical/imaging data provided
Different treating philosophies
Insufficiently stringent guidelines e.g. target volume selection
Deliberate deviation from guidelines which are perceived as too radically different from standard practice
Insufficient understanding/application of ICRU
Differences reflect areas of controversy with no globally accepted standard practice
Institutional differences in treatment planning e.g. margins allowed for set-up reproducibility; modelling of build-up regions; calculation algorithms
Deviations introduced in the course of importing or exporting a digital dataset

Table 1.3: Proposed reasons for non-adherence to *dummy run* guidelines. Reprinted from A. Fairchild et al. **Do results of the EORTC dummy run predict quality of radiotherapy delivered within multicentre clinical trials?** *European Journal of Cancer* 48 (2012), pp. 3232–3239, © 2012, with permission from Elsevier.

A landmark EORTC trial including an exhaustive RTQA programme for that time was the EORTC 22922/10925 study. Four steps were within this specific RTQA programme:

1. Verification of the consistency of the data at the EORTC Data Centre by performing a double data entry procedure in the centralised database.

2. Completion of a *dummy run*.

3. Early reviewing of patient eligibility and treatment compliance.

4. Mailing of thermoluminescent dosimeters to evaluate electron dosimetry.

When evaluating the results of the *dummy run*, significant deviations from protocol were found for treatment setup and prescription, causing a significant variation of the prescribed dose to the internal mammary chain - medial supraclavicular fossa in 10% of the cases for the control group, and in 21% for the nodal-irradiation group. These deviations would likely affect the survival benefit this study was powered to detect, which would decrease to a true 3.8% benefit as opposed to the initially calculated 5% benefit to be found with nodal-irradiation. Thus, this could potentially lead to a false negative result of the study. Looking on the bright side, since the *dummy run* was performed early in the trial, this allowed recommendations to be made to centres based on the results to avoid the deviations in future patients. Also, the protocol was amended based on these recommendations (Poortmans et al., 2001). Ultimately, 10 year overall survival results proved, in effect, to be marginally, but not significantly better in the nodal-irradiation arm (82.3%) vs. 80.7% in the control arm (HR for death with nodal irradiation, 0.87; 95% CI, 0.76 to 1.00; $p\text{-}value = 0.06$) (Poortmans et al., 2015).

In 2006 the EORTC RTQA platform established five levels of RTQA requirements for centres participating in EORTC trials **Table 1.4** . Level 1 includes general strategies to ensure minimal acceptable quality radiotherapy to be administered in all participating centres. Further levels are protocol-specific credentialing, to be adapted depending on the aim of the trial and the technological complexity of the radiotherapy techniques (Weber et al., 2011).

Level 1	• **Facility questionnaire**
	• **External reference dosimetry audit**
Level 2	• **Dummy run**
Level 3	• **Limited individual case Review**
Level 4	• **Extensive individual case review**
Level 5	• **Complex dosimetry check**

Table 1.4: EORTC levels of RTQA. Reprinted from D. C. Weber et al. **Quality assurance for prospective EORTC radiation oncology trials: The challenges of advanced technology in a multicenter international setting**. 2011, © 2011, with permission from Elsevier.

RTQA centres guarantee homogeneous radiotherapy treatment delivery between sites included in clinical trials, ameliorating protocol compliance, minimising deviations from protocol, revealing systematic errors during the clinical treatment process and detecting misinterpretations or misunderstandings of trial protocols. RTQA plays an important role in clinical trials, even when radiotherapy does not constitute the main study endpoint, although its intensity may vary depending on the weight radiotherapy has on the study question. Thus, all RTQA programmes need to pursue treatment homogeneity to obtain a trial compliant interpretable population without hindering patient accrual (FitzGerald, 2012). Unfortunately, some RTQA programmes demanded many human, technical and monetary resources, ultimately hampering trial accrual. This, and the fear of losing participating centres which are not compliant with the RTQA process, has caused reluctancy from study coordinators, participating centres and/or cooperative groups to participate in obligatory RTQA credentialing (Weber et al., 2012). Another issue is that results of trials with excellent RTQA may not correspond with results from standard- RTQA used in clinical practice (Bekelman et al., 2012). These issues highlight the need for a rational design of RTQA programmes, which should be well adapted to the requirements of each particular clinical trial.

Once the key role of RTQA within clinical trials was clearly established, early on, RTQA guidelines were developed by societies such as European Society for Radiotherapy & Oncology (ESTRO) (Thwaitesa et al., 1995). Nowadays specific RTQA procedures are available, the main ones are summarised in a table including all centres in the Global Clinical Trials RTQA Harmonization Group (GHG) (**Table 1.5**). They allow for a quite rigorous quality assurance process, but unfortunately not all centres use the same nomenclature, which

complicates the intercomparison of these procedures. To remedy this situation, the GHG has made an important effort to reach a common language, specified in **Table 1.6**, where they were able to produce a homogeneous set of 10 out of the 27 RTQA procedures analysed from the GHG centres (Melidis et al., 2014). Even templates, which can be adapted to trial situations, have been written to facilitate the preparation of RTQA programmes within clinical trials (Nilsson et al., 2014).

A final and quite necessary step to facilitate the logistics of RTQA procedures was the development of quality control platforms for digital data transfer (Fairchild et al., 2012b). These platforms have considerably evolved over the past two decades, leading to real time quality control servers where participating centres may upload their data, allowing it to be reviewed by the central reviewers. As previously mentioned, RTQA requirements have much evolved due to the complexity of modern radiotherapy techniques, calling for exponentially increasing amounts of data which need to be received in an accurate and timely fashion to allow interpretation by central RTQA reviewers. A digital data facility allows prospective collection of RTQA data with real-time review of ICR, facilitating early detection of trial deviations, leading to a prompt implementation of corrective actions or protocol amendments if necessary. Another notable advantage is the possibility of performing real-time case reviews of contouring and dosimetry procedures before a trial patient is treated, thus ensuring treatments are performed per protocol (Fairchild et al., 2012b; Willett et al., 2012). Digital data transfer facilities also reduce RTQA costs by eliminating mail and storage costs, site visits and reducing review time (digital data is quicker to revise than hard copies). The EORTC's response to this need is the Visualization and Organization of Data for Cancer Analysis (VODCA) programme. It includes digital radiotherapy imaging, treatment planning and verification data for review. The VODCA platform provides quick, thorough, prospective patient reviewing for international multicentre clinical trials, and can assist multinational or intergroup collaborations. Currently, a team is necessary for collection and verification of the data, followed by its review. This is especially needed since collecting data through digital data transfer quality control platforms often deals with the issue of the homogeneity of the nomenclature used for region of interest (ROI), DVH, toxicity, dose and volume units. In an endeavour to standardise this nomenclature, the American Association of Physicists in Medicine (AAPM) has launched the task group 263 (Mayo et al., 2015). This is a barrier which will need to be surpassed to advance towards future directions such as the integration of automatic review software, which would much relieve central reviewers of this systematic part of RTQA, allowing them to focus on other aspects. Another very promising future in-

	Baseline	Preaccrual		Prospective/retrospective RTQA data capture and analysis	
Member	Facility and basic dosimetric requirements	Protocol-compliant dummy patient or site connectivity check	Advanced RT technique credentialing	Remote review of patients' RT treatments	Site visits
ATC-ITC	Not performed	Data submission test or rapid review or dry run	Not performed	1. Basic archiving 2. 1 + completeness check 3. 2 + compute DVHs 4. 3 + reconcile structures 5. 4 + image registration and case report forms	Not performed
EORTC-ROG	Facility questionnaire and external reference dosimetry audit by RPC or other QA office	Dummy run or digital data integrity quality assurance	Complex dosimetry check or virtual phantom procedure	Individual case review and case report forms	Not performed
IAEA	Facility questionnaire and reference beam output	Benchmark cases	Complex dosimetry check	Individual case review and case report forms	Not performed
JCOG	Facility questionnaire and external reference dosimetry audit	Dummy run or digital data integrity quality assurance	Complex dosimetry check	Individual case review and case report forms	Not performed
RPC	Facility questionnaire and OSLD/TLD beam output audit	Benchmark cases or rapid review	Credentialing for advanced technology clinical trials or complex dosimetry check	Review of patients' treatment records, timely reviews, and case report forms	On-site dosimetry review visits
RTOG	Facility questionnaire and external reference dosimetry audit by RPC	Dummy run or rapid review or dry run	Credentialing for advanced technology clinical trials	Individual case review and case report forms	On-site dosimetry review visits by RPC
RTTQA	Facility questionnaire and external reference dosimetry audit by RTTQA or IPEM/NPL*	Outlining and planning cases/exercises or pretrial case review	Credentialing for advanced techniques	Individual case review and case report forms/plan assessment form	On-site dosimetry visit
TROG	Facility questionnaire and external reference dosimetry audit (eg, ARPANSA/ACDS)	Benchmark cases or dummy run or digital data integrity quality assurance	Credentialing for advanced technology clinical trials	Individual case review and case report forms	On-site dosimetry review visits by TROG (or approved service)

Abbreviations: ATC = Advanced Technology Consortium; ARPANSA/ACDS = Australian Radiation Protection and Nuclear Safety Agency/ Australian Clinical Dosimetry Service; EORTC-ROG = European Organization for the Research and Treatment of Cancer—Radiation Oncology Group; GHG = Global Clinical Trials RTQA Harmonization Group; IAEA = International Atomic Energy Agency; IPEM/NPL = Institute of Physics and Engineering in Medicine/National Physical Laboratory; ITC= Image-guided Therapy QA Center; JCOG = Japan Clinical Oncology Group; OSLD/ TLD = optically stimulated/thermoluminescent dosimeter; RPC = Radiological Physics Center; RTOG = Radiation Therapy Oncology Group; RTQA = Radiation Therapy Quality Assurance; RTTQA = Radiation Therapy Trials Quality Assurance; TROG = TransTasman Radiation Oncology Group.

Table 1.5: RTQA procedures used by the members of the GHG steering committee. Reprinted from C. Melidis et al. **Global Harmonization of Quality Assurance Naming Conventions in Radiation Therapy Clinical Trials**. *International Journal of Radiation Oncology Biology Physics* 90.5 (2014), pp. 1242–1249, © 2014, with permission from Elsevier.

novation may be its integration with the platforms of headquarter imaging and translational research units (Fairchild et al., 2012b).

Lastly, study reviews and a meta-analysis have found an association between RTQA and clinical patient outcomes. A review including nine trials (Weber et al., 2012) presented

Naming convention categories	Current name	New name
Baseline	Facility questionnaire	Facility questionnaire
	External reference dosimetry audit	Beam output audit
	OSLD/TLD beam output audit	
	Reference beam output	
Preaccrual	Benchmark case by RPC and TROG	Benchmark case
	Dry run by ITC	
	Dummy run by EORTC-ROG, JCOG, and RTOG	
	Outlining and planning cases/exercises by RTTQA	
	Digital data integrity quality assurance	Dummy run (without delineation exercise)
	Dry run by RTOG	
	Data submission test	
	Benchmark case by EORTC-ROG, IAEA	Dummy run (with delineation exercise)
	Dummy run by TROG	
	Pretrial case review	
	Rapid review	
	Credentialing for advanced technology clinical trials	Complex treatment dosimetry check
	Complex dosimetry check	
	Credentialing for advanced techniques	
	Virtual phantom procedure	Virtual phantom
Prospective/retrospective RTQA data capture and analysis	Image registration by Advanced Technology Consortium (having completeness check, basic archiving, compute dose—volume histograms, and reconcile structures as prerequisites)	(Prospective or retrospective) individual case review
	Individual case review	
	Timely review	
	Review of patients' treatment records	Review of patients' treatment records
	Case report forms	Case report forms
	On-site dosimetry review visit by RPC	Protocol compliance and dosimetry site visit
	On-site dosimetry review visit by TROG	
	On-site dosimetry visit	

Table 1.6: Procedures described by the naming conventions of RTQA procedures agreed upon by the members of the GHG steering committee. Abbreviations as in **Table 1.5**. Reprinted from C. Melidis et al. **Global Harmonization of Quality Assurance Naming Conventions in Radiation Therapy Clinical Trials**. *International Journal of Radiation Oncology Biology Physics* 90.5 (2014), pp. 1242–1249, © 2014, with permission from Elsevier.

the results of the RTQA programme with the corresponding patient outcome for six trials (**Table 1.7**). They found that RTQA deviations affected the primary study aim in 62.5% of the nine reviewed trials. Fairchild et al., (2013) have reviewed the evidence supporting the correlation of RT quality with clinical outcomes in seventeen multicentre clinical trials and one Patterns of Care Study (Fairchild et al., 2013). Seven showed significantly higher failure rates after inadequate versus adequate RT. Five of nine and two of five studies reported significantly worse overall and progression-free survival after poor quality RT, respectively. They conclude that protocol-compliant RT may decrease failure rates and increase overall survival and probably affects the ability to answer the central trial question. A meta-analysis performed earlier the same year, including eight cooperative group trials, reached a similar conclusion, as radiotherapy protocol deviations were significantly associated with higher risks of failure of the treatment and overall mortality (Ohri et al., 2013).

Study [ref]	Type of QA	Number of cases evaluated *n* (%)	Minor deviations *n* (%)	Major deviations *n* (%)	Technical issues with QA review *n* (%)	Impact on clinical outcome	*p* Value
HD 4 [5]	R	368 (98.0)	–	141 (37.5)*	8 (2.1)	7-year RFS with D: 72% *vs.* 7-year RFS with no D: 84%	0.004
EORTC 20884 [2]	R	135 (88.8)	–	63 (46.7)	46 (30.3)	5-year RFS with D: 90% *vs.* 5-year RFS without D: 84%	0.31
RTOG 0411 [4]	R	NS	–	13 (13.4)	NS	Grade GI ⩾ 3 toxicity with D:45%‡ *vs.* Grade GI ⩾ 3 toxicity without D:18%‡	0.05
RTOG 9704 [1]	R	416 (92.2)	–	200 (48.0)**	14/35 (40.0)†	mOS with D: 1.46 yo *vs.* mOS without D: 1.74 yo	0.008
RTOG 0022 [8]	R	67 (97.0)	47 (89.0)	6(11.0)	14/67 (21.0)	LRF with major D: 50% *vs.* LRF with no major D: 6%	0.04
TROG 0202 [15]	P & R††	687 (80.5)‡‡	–	97 (11.8)	33/820 (4.0)	OS with major D: 70% *vs.* OS without major D: 50%	<0.001

Abbreviations: R, retrospective; P, prospective; LRF, local-regional failures; D, deviations; mOS, median overall survival; RFS, relapse-free survival; GI, gastro-intestinal; NS, not specified.
* Deviations were scored as adherence per protocol or less than per protocol (see main text for details).
** Deviations were scored as adherence to protocol-defined volumes, dosimetry, treatment time and technical delivery characteristics. (see main text for details).
† Denominator is influenced by the number of patients with negative clinical outcome and/or the absence of delivered RT (see main text for details).
†† QA of the TROG study was performed with a primary (interventional) prospective review and secondary review. Figures provided in Table are from the second retrospective review.
‡ Observed toxicity during chemo-radiotherapy.
‡‡ Number of evaluated cases in the interventional prospective QA program.

Table 1.7: Correlation between RTQA programmes and patient outcomes in 6 prospective clinical trials. Reprinted from D. C. Weber et al. **QA makes a clinical trial stronger: Evidence-based medicine in radiation therapy**. *Radiotherapy and Oncology* 105.1 (2012), pp. 4–8, © 2012, with permission from Elsevier.

1.3.1 The importance of adequate delineation of regions of interest

Delineation of treatment plans is key in radiotherapy treatment. Proper delineation of both TV and OAR allows for an optimal oncological treatment along with more precise knowledge of the dose administered to the surrounding healthy tissue. Due to this fact, several studies have evaluated both the interobserver and the intraobserver variability between contours (Rasch et al., 1999; Petric et al., 2008; van Mourik et al., 2010; Petrič et al., 2013).

Since the 1990's studies presenting results from RTQA programmes within multicentre trials have found that variability in the delineation of TV was the factor with the greatest impact on RTQA deviations (Kouloulias et al., 2004). And not only is this the case for TV definition. An analysis of the impact of variations in rectal contouring, when considering interobserver variability, within a multicentre trial on prostate cancer showed that even when most observers agreed on rectal definition, contouring was still an important source of uncertainty (Foppiano et al., 2003). To evaluate whether the use of protocol guidelines for contouring decreased interobserver variations, a contouring study for postoperative radiotherapy in lung cancer was performed within the international multicentre Lung Adjuvant Radiotherapy Trial (LungART) (Spoelstra et al., 2010). Ten physicians sent in their routine clinical target volume (CTV) (per their routine clinical practice) and six sent in routine and protocol CTV for two cases. There was a variation of up to three times between physicians,

which decreased significantly with the use of the LungART protocol (Spoelstra et al., 2010). These results support the need to perform obligatory *dummy runs* within multicentre trials in the 3D conformal radiotherapy era.

To highlight this role of ROI delineation within RTQA, pertinent results were found in a study performed within the RTOG which performed a detailed review of radiotherapy guideline compliance within recent RTOG digestive tract trials. Components of the radiotherapy process were scored and ultimately deemed as according to per-protocol, variation acceptable or deviation unacceptable. When scoring anal cancer trials (three phase III trials: RTOG 9405, RTOG 9704, and RTOG 9811 and an IMRT trial: RTOG 0529), most unacceptable deviations were due either to errors in field design or in contouring of the GTV, and very rarely due to errors in dose or fractionation (Willett et al., 2012).

1.3.2 Application in cervical cancer

Specifically, in the field of cervical cancer, with the implementation of modern day radiotherapy techniques, such as IMRT and IGBT, RTQA acquires an even higher significance. In recent years, the use of these advances in EBRT, and BT (especially IGBT), have yielded excellent 3 year local control rates (92% in tumours >5 cm and 98% in tumours 2-5 cm) (Pötter et al., 2011). The achievement of these rates was by the use of Gyn GEC-ESTRO recommendations for contouring of the HR-CTV and for evaluation of dose volume constraints for OAR (Haie-Meder et al., 2005).

Results for RTQA in LACC cervical cancer were presented for the Gynecologic Oncology Group (GOG) protocol 165 in an abstract. They compared retrospectively high dose rate (HDR) BT treatments administered in centres that were fully credentialed before including patients in the trial with those that were administered in non-credentialed centres. All HDR BT treatments delivered in credentialed centres had no major protocol deviations, while of those delivered in non-credentialed centres only 81% had no major deviations (Lowenstein et al., 2002).

One of the first fully-published trials detailing RTQA in cervical cancer was for early stages (Toita et al., 2009). This Japanese multicentre prospective trial (JAROG0401/JROSG04-2) evaluated the outcomes and side effects of exclusive radiotherapy, including HDR BT administration. All the treatment plans of the 60 patients included in this trial had an ICR. Modern radiotherapy techniques such as IMRT and IGBT were not used in these patients, and EBRT fields were evaluated per 2D bony landmarks. The data required for

the ICR were baseline MRI, EBRT treatment charts, digitally reconstructed radiographs or simulation films, electronic portal imaging device dose distributions or verification portal films, HDR BT treatment charts, films and isodose distributions for all applications. Of all deviations, the most frequent one was non-compliance for point A definition, followed by differences between simulation and verification films of over 5 mm. In general there was good protocol compliance within this study, although the lesser complexity of the techniques most probably favoured these results (Toita et al., 2009).

More recently, more rigorous RTQA has been performed in the field of cervical cancer, within the EMBRACE and the Induction Chemotherapy Plus Chemoradiation as First Line Treatment for Locally Advanced Cervical Cancer (INTERLACE) trials (ClinicalTrials.gov NCT01566240, 2012). The RTQA in these trials has evolved in response to the more complex radiotherapy techniques included in these studies, such as IMRT (optionally) or IGBT (optional for the latter).

The importance of adequate delineation of regions of interest in cervical cancer

Little has been published concerning uncertainty in EBRT TV contouring for cervical cancer (Weiss et al., 2003; Wu et al., 2005). In one study, seven clinicians (five radiation oncologists and two gynaecologists) contoured all CTV for three cervical cancer patients. The inclusion of the adequate regions in each CTV was fairly consistent amongst observers. However, there was flagrant interobserver variability. They concluded that CTV delineation seemed to influence the accuracy of treatment delivery more than other factors, such as patient set-up or organ motion (Weiss et al., 2003).

Such evidence led to a quite exhaustive RTQA programme performed within the INTERLACE trial (Eminowicz et al., 2016a; Eminowicz and Mccormack, 2015; Eminowicz et al., 2016b). Although none of the participating trial centres were using IMRT at the time, in the perspective of implementing this technique for LACC treatments in participating centres, a detailed atlas was developed to aid delineation (Eminowicz et al., 2016a). Their aim was to reduce interobserver contouring variability, as it has been shown that guideline use reduces this variability in other tumour types. The RTQA team performed a literature search for guidelines for ROI determination in cervical cancer, and after identification of seven papers they found eleven areas of variation. The bowel, femur, vagina, parametria, nodal borders including cranially and caudally, para-aortic nodes, and the margin around macroscopically enlarged nodes were within these areas. This information was used to create consensus con-

touring guidelines for the INTERLACE protocol, and the proportion of protocol-compliant contours before and after the implementation of these guidelines was analysed. Four ROI were analysed to this end, primary CTV, CTV-nodes, bladder and rectum. The average proportion of protocol compliant contours increased by 1.8, then to 2.7 after atlas implementation (difference of 0.9; 95% CI, 0.3-1.5; $p\text{-}value = 0.003$) , clearly reducing interobserver differences in contouring.

Considering the excellent, previously mentioned, local control results obtained with IGBT, this breakthrough in treatment for cervical cancer has prompted several studies, culminating in the EMBRACE initiative. Some of these studies have found that detailed guidelines and training in contouring can improve consistency in GTV and CTV definition in IGBT (Petric et al., 2008; Dimopoulos et al., 2009)

However, little data have been published on specific IGBT RTQA. To date, the most representative data comes from the EMBRACE trial (Kirisits et al., 2015). In this study there were no specifications for BT dose prescription, centres were to treat according to their institutional protocols (for dose, fractionation, dose rate and technique). Considering that centres had different BT techniques (applicator types, type of imaging and planning used), they did not all receive the same clinical case to perform the *dummy run*, rather they were asked to send in two of their own cases which met certain clinical requirements. Only BT contouring of TV, dose and volume reporting was required to be per Gyn GEC-ESTRO recommendations. EBRT was administered per EMBRACE guidelines for TV definition, planning, dose and fractionation, and overall treatment time. Nine of the 30 participating centres were considered as experienced in IGBT. For both EBRT and BT techniques, the largest variations were due to contouring. As expected, there were less protocol deviations for experienced centres with respect to those with less experience in IGBT (Kirisits et al., 2015).

All of this evidence supported the performance of a *dummy run* within the RAIDs trial, to ensure that the patients identified as non-responders to standard chemoradiation treatment were in effect resistant to treatment, and had not simply received inadequate chemoradiation. This way correctly identified non-responders could be correlated with potentially treatable genomic, proteomic, biochemical, molecular, virological or cellular biology alterations.

As the RAIDs trial included very geographically distant countries, a practical approach was to perform this *dummy run* online. Mcenery et al., (1995) found that online educational tools offer the possibility to achieve interactive medical instruction for geographically dispersed students. However, especially in the medical field, it can sometimes be difficult

for both the faculty and health care professionals to use the technology efficiently and effectively. Experience within this field seems to suggest that an optimal model of delivery could include both face-to-face and e-learning components, deemed as *blended learning* (Sharpe et al., 2006).

Flexibility is a basic consideration within e-learning programmes, especially within highly demanding professions such as those found in the health-care setting. A systematic review has found that this is often manifested as *learner control*, which demands different entry points and learning trajectories during a course, with self-regulation of task management and adaptation to local and personal circumstances (Booth et al., 2009).

An important issue in e-learning is student outcome evaluation. Very few reports include objective internal testing to validate web-based learning tools as a primary outcome (Kronz et al., 2000; Erickson et al., 2003; Ridgway et al., 2007).

The purpose of this study is to validate the methodology of this new concept of online delineation workshop (ODW) in LACC within the European multicentre prospective study RAIDs, by reviewing the participant contours in the different stages as well as their personal perception of the knowledge acquired.

2
Aims

Contents

2.1 Primary endpoints:

1. Interobserver validation of the online delineation workshop (ODW) in locally advanced cervical cancer (LACC) by using qualitative and quantitative assessments of the improvement of the clinician contours between clinicians (interobserver variability).

2. Intraobserver validation of the ODW in LACC by assessing qualitatively and quantitatively the improvement of the clinician contours for each clinician (intraobserver variability).

2.2 Secondary endpoints:

1. Evaluation of the teaching methodology of the online contouring workshop as reflected by the participant satisfaction questionnaires.

2. Analysis of clinician contouring on magnetic resonance imaging (MRI) for European centres which use MRI for brachytherapy planning versus those which do not.

3
Materials and methods

Contents

3.1 Assessment of locally advanced cervical cancer radiotherapy treatments in RAIDs centres

The first step was to assess general practice in participating clinical Rational molecular Assessments and Innovative Drug Selection (RAIDs) centres. For this purpose, a general questionnaire about locally advanced cervical cancer (LACC) radiotherapy was prepared and sent to RAIDs centres (**Table 3.1**). This questionnaire allowed for a basic baseline evaluation of their local treatment protocols before their enrollment in the online delineation workshop (ODW). The RAIDs trial was observational and collected different treatments administered to patients through the electronic case report forms (eCRFs), but did not require following any particular treatment protocol, other than radiotherapy contouring guidelines specified in the ODW both presented interactively and supplied in writing. Most interestingly, this questionnaire allowed the evaluation of the experience of the centres in the treatment of cervical cancer, as well as the techniques used for external beam radiotherapy (EBRT) and brachytherapy (BT).

Which best describes your institution?
A) Private centre
B) Academic centre
C) Public centre
How many cervix patients do you approximately treat per year with definitive radiotherapy including external beam therapy, chemotherapy and brachytherapy?
A) 0-50
B) 50-100
C) $\geq$ 100
Which cervix cancer patients do you treat with definitive radio(chemotherapy)?
A) Patients with positive pelvic and/or paraaortic lymph nodes
B) Patients with positive paraaortic lymph nodes
C) If distant metastasis is excluded I treat all patients with tumour stages I-IVa
D) Patients with negative lymph nodes if local tumour stage is IIb or greater
Which applicator do you use?
A) Tandem/ring
B) Ovoids
C) Mould
D) Tandem/cylinder
E) Interstitial needles
Do you use interstitial application techniques for brachytherapy?
A) Yes
B) No
Which dose rate do you use?
A) high dose rate (HDR)
B) pulsed dose rate (PDR)
C) low dose rate (LDR)
Which kind of imaging do you perform with the applicator in place at the time of brachytherapy?
A) X-ray
B) C-arm cone-beam CT (CBCT)
C) computed tomography (CT)
D) magnetic resonance imaging (MRI)
E) ultrasound (US)
In case you are not doing 3D image based dose planning: when do you plan to start?
A) in the next 6 months
B) within one year
C) within 3 years
D) I don't know / don't plan to
To which volume/point do you prescribe brachytherapy dose?
A) Point A
B) high risk CTV (HR-CTV)
C) intermediate risk CTV (IR-CTV)
D) Other
Which total EQD2 dose do you prescribe to the target volume/point indicated above?
A) <65 Gy
B) 65-74 Gy
C) 75-79 Gy
D) 80-84 Gy
E) $\geq$ 85 Gy
Number of brachytherapy fractions
A) 1 or 2
B) 3
C) 4
D) 5
E) 6
Which dose volume constraints are used for organs at risk?
A) Bladder
B) Rectum
C) Sigmoid
D) Vagina

Table 3.1: Pre-workshop questionnaire.

3.2 Online delineation workshop structure

The number of participants needed to find statistically significant differences (number needed to treat (NNT)) for inter and intraobserver contouring variability was calculated, resulting in 50 participants. Thus, two to four participants were invited from each centre (proportional to the gynaecological team) to participate in an ODW in LACC. The goal was an estimated participation from the 22 RAIDs clinical centres of approximately 50 clinicians **Table 1.2**. This exceeded the capacity of the ODW, so two identical workshops were planned.

Between two and four participants from each centre were enrolled in an ODW in LACC. A technical partnership was agreed upon with the European Society for Radiotherapy & Oncology (ESTRO) and the methodology resembled the one used in Fellowship in Anatomical deLineation and CONtouring (FALCON) ESTRO ODW. The live presentations were performed via WebEx conferencing and the contouring was on the FALCON EduCase^{TM} online contouring platform.

Training was provided by an expert in the field, CHM, and for every 10 participants there was one tutor. The tutors were radiation oncologists with experience in the field of cervical cancer, already trained to use the FALCON EduCase^{TM} online contouring system. Tutors gave the participating clinicians support to assist them with the use of FALCON EduCase^{TM}. They also assisted by answering basic questions about the use of the contouring guidelines, and compiled more specific contouring questions to be answered by the expert, CHM, during the live WebEx sessions. The live sessions were completed in 3 weeks and the participants delineated both the EBRT treatment on computed tomography (CT) as well as the subsequent BT treatment for the same clinical case. The BT treatment was contoured on magnetic resonance imaging (MRI). The case and the image sets for both treatments were reviewed and chosen from the ESTRO FALCON EduCase^{TM} contouring library with the expert, CHM.

The live online workshop sessions for the first ODW were held on June 20^{th}, June 26^{th} and July 10^{th}, 2013, and the sessions for the second ODW were held on January 7^{th}, 13^{th}, and 27^{th}, 2014. The first two live online sessions of each workshop were presented by the tutors.

Session 1: In the first presentation they explained how to use the FALCON EduCase^{TM} contouring platform. The participants were also explicitly informed (orally and in writing within the presentation slides), that their contours would be used for a study

evaluating the ODW and their conformity was requested. No participant revoked their conformity. The clinicians had a six day period to perform baseline contouring (C1) (how they contour in daily practice).

Session 2: Contouring guidelines for both EBRT and BT following the An intErnational study on MRI guided BRachytherapy in locally Advaced CErvical cancer (EMBRACE) protocol were presented and C1 was reviewed, followed by a question-and-answer session. Afterwards, the recommendations from the Groupe Européen de Curiethérapie – European Society for Radiotherapy and Oncology (Gyn GEC-ESTRO) working group, EMBRACE[1] protocol, a pelvic nodal atlas and two consensus atlases for contouring of normal tissue in the pelvis were sent to the clinicians to further aid delineation (Haie-Meder et al., 2005; Taylor et al., 2005; Gay et al., 2012; Gay et al., 2011). Clinicians were given two weeks to modify their previous contours for the same two image sets. This was considered the guideline contouring (C2).

Session 3: In the 3rd week the expert in the field, CHM, reviewed C1 and C2 and there was a question-and-answer session.

During the whole ODW period tutors contacted the clinicians and answered their questions, and the participants were supplied with links to recordings of each live session as well as the corresponding presentation slides.

Lastly, clinicians recontoured both EBRT and BT treatments after seeing the reference contours and having their questions answered. This final contouring (C3) was done between 1.5-2 months after the last live presentation (session 3), so as to evaluate the long term teaching impact.

3.3 Description of the clinical case

The 45 year old patient was diagnosed with a FIGO IIIB squamous cell carcinoma of the cervix. The gynaecological exam revealed a large exo-endophytic growth (85x50x60 mm) involving the vagina (all fornices 1 cm, anterior wall 4 cm). The tumour infiltrated the right parametrium proximally and the left one up to the pelvic side wall (**Figure 3.1**, **Figure 3.2**). There was no involvement of bladder mucosa (cystoscopy). Abdominopelvic CT showed a cervical enhancing mass with vaginal involvement, enlarged external, internal,

[1] http://www.embracestudy.dk

lower common iliac, and pre-sacral nodes, without paraaortic nodes. The response to EBRT and concurrent chemotherapy was good, with resulting tumour dimensions of 55x40x30 mm, free right parametrium, induration of half of the left parametrium, and involvement of 1cm of the anterior vaginal wall at the time of BT (**Figure 3.3**, **Figure 3.4**).

Compulsory volumes for contouring exercises (at minimum required slices for organs at risk (OAR) and whole region of interest (ROI) for target volume (TV):

- Case 1 - EBRT:
  - OAR: Bladder, rectum, bowel, sigmoid.
  - cervix, parametria and vaginal gross disease (GTV-p).
  - nodal elective volume (CTV-node).
  - radiologically pathological lymph nodes to boost (GTV-node).
  - GTV-p, uterus and vagina, at least 20 mm below GTV-p (CTV-p).
- Case 2 - BT:
  - OAR: Bladder, rectum, sigmoid.
  - gross tumor volume (GTV): Macroscopic tumour (if present) at time of BT.
  - high risk CTV (HR-CTV): Macroscopic tumour extension at time of BT + whole cervix + presumed extra cervical tumour extension.
  - intermediate risk CTV (IR-CTV): HR-CTV + macroscopic tumour extension at diagnosis providing a minimal margin of 10 mm to residual disease at time of BT in direction of potential spread.

3.4 Contour evaluation methodology

Intraobserver variability of each participant was evaluated before and after the presentation of contouring guidelines (C2 vs. C1), before and after the expert review of contours (C3 vs. C2) and the final contour compared to baseline (C3 vs. C1) on one advanced stage cervical cervical cancer case, for both EBRT and BT treatments. Quantitative and qualitative differences were assessed. Interobserver variability was determined quantitatively by analyses based on ROI, years of experience in the field, and for BT also between centres using MRI-based image guided brachytherapy (IGBT) and others.

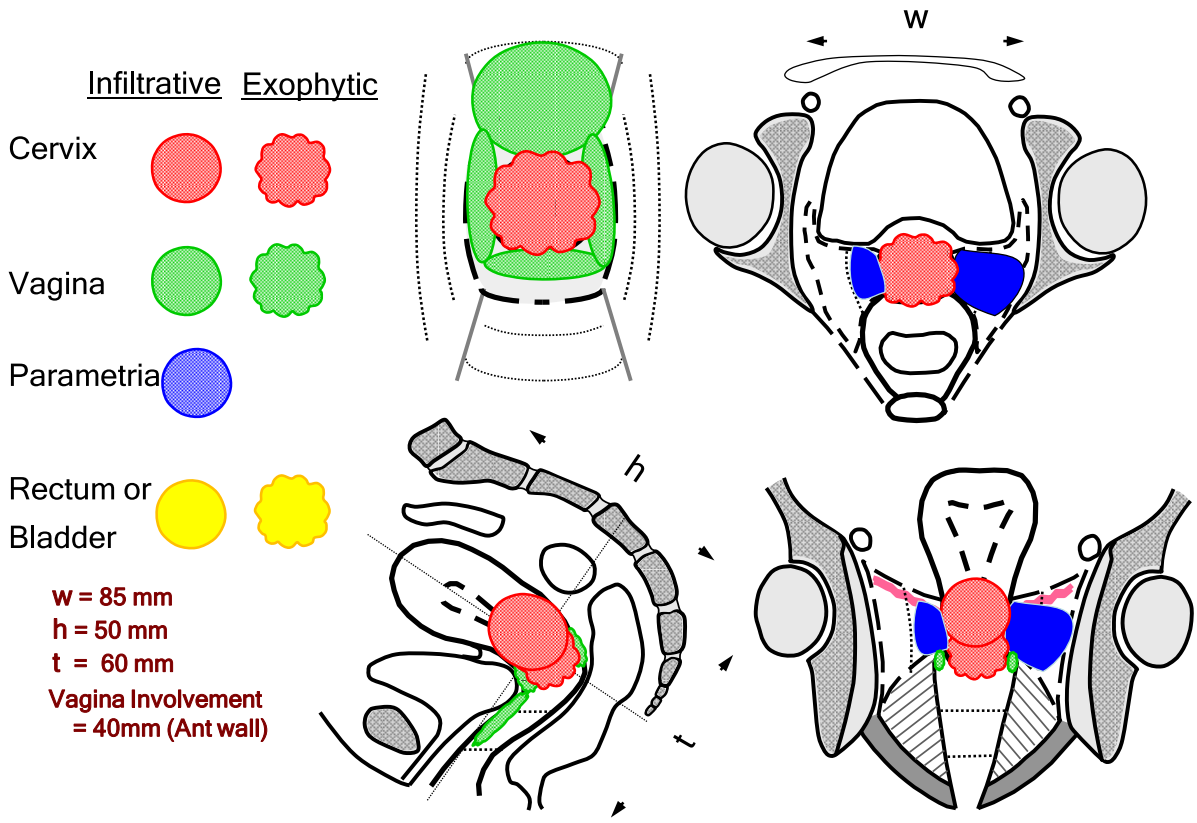


Figure 3.1: Initial clinical drawing at diagnosis, based on the clinical gynaecological exam.

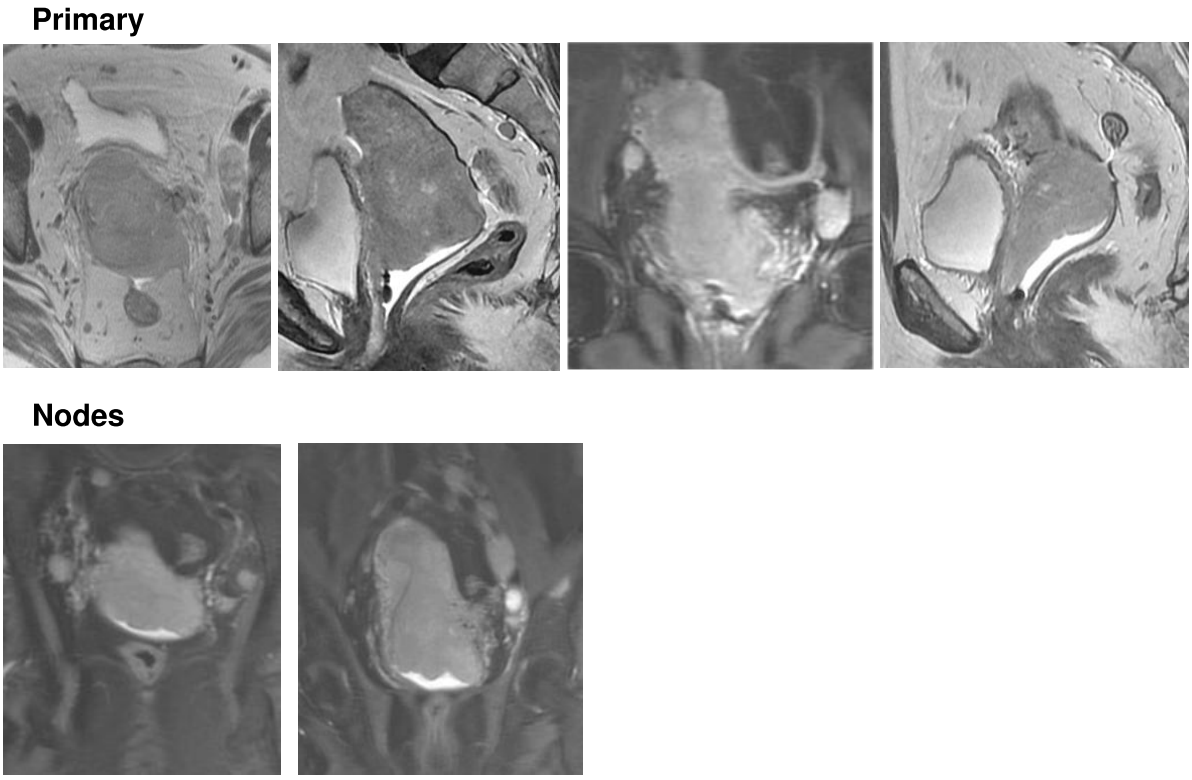


Figure 3.2: Baseline MRI findings at diagnosis.

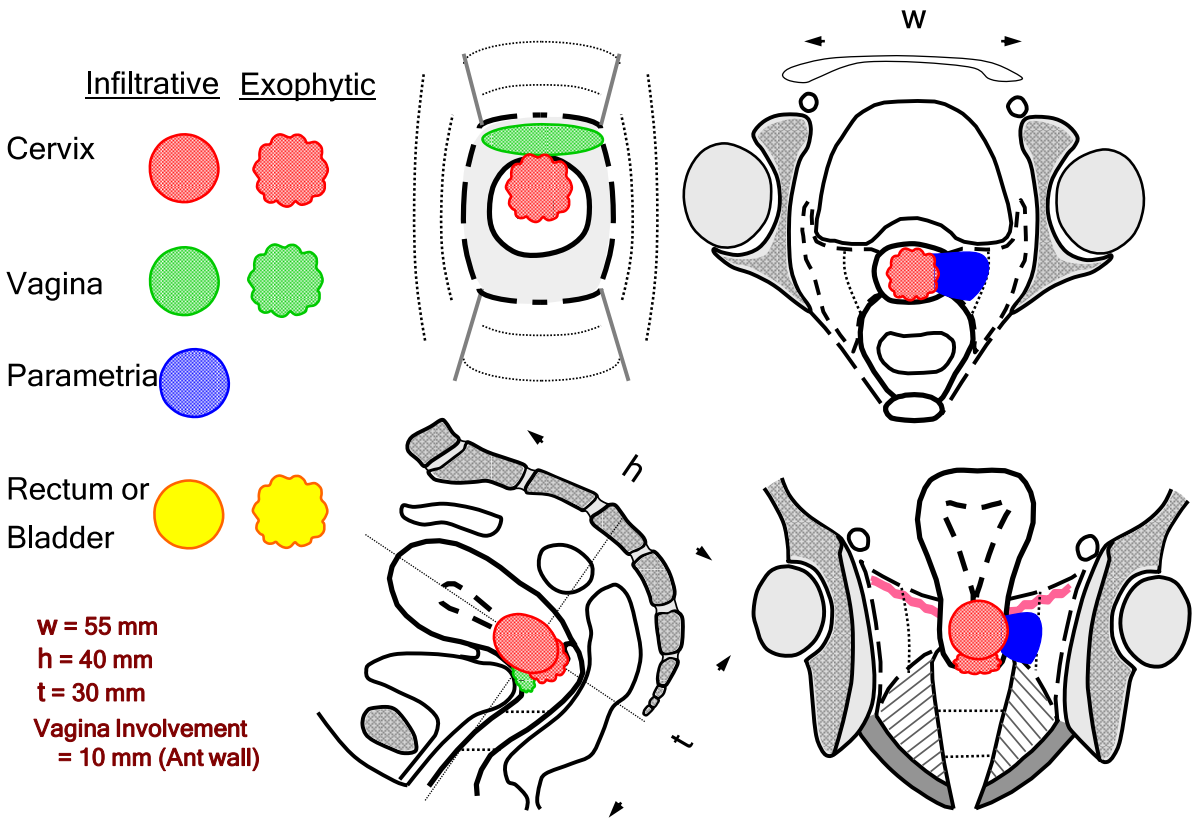


Figure 3.3: Clinical drawing at the time of BT, based on the clinical gynaecological exam.

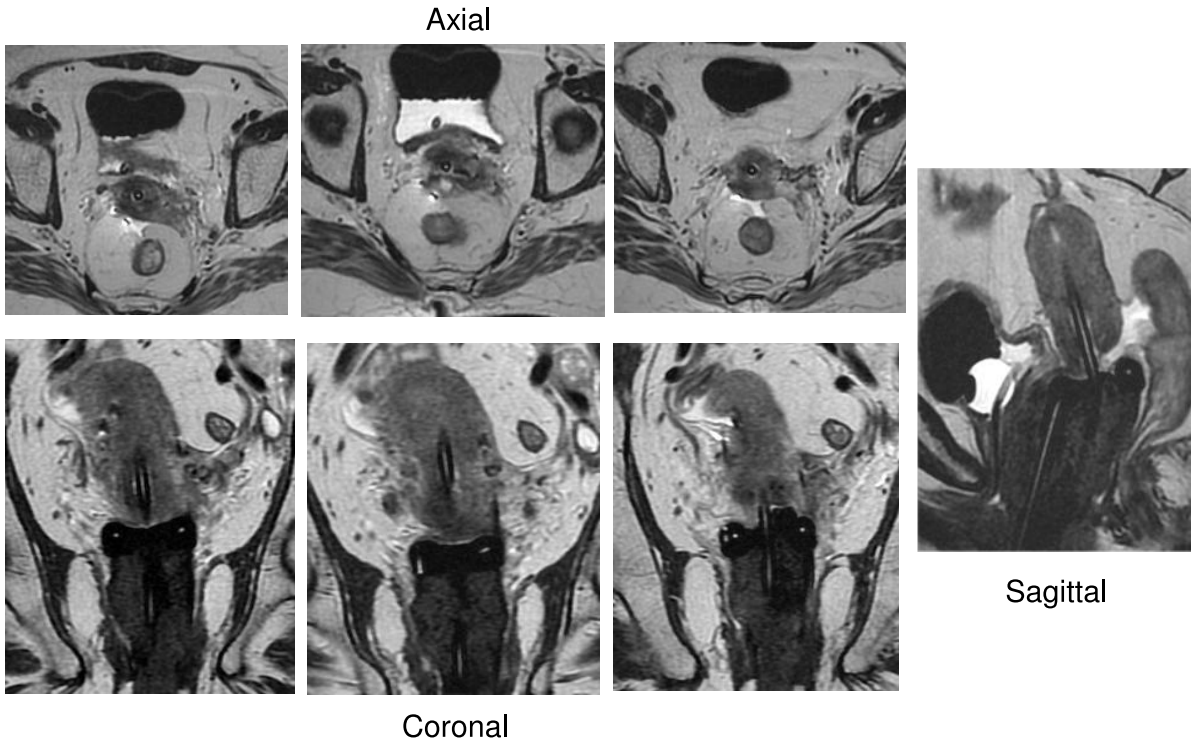


Figure 3.4: MRI images at the time of BT.

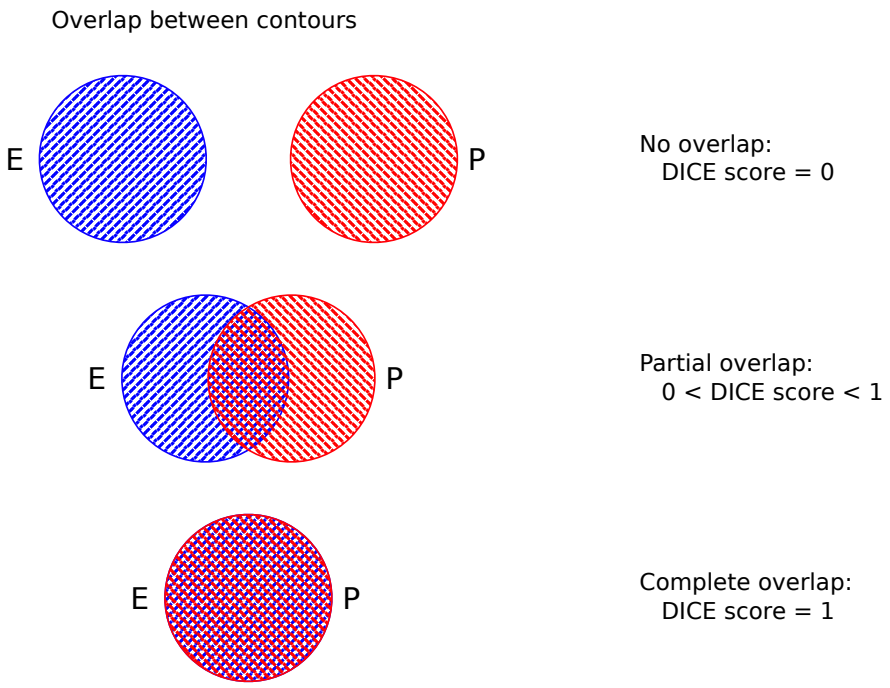


Figure 3.5: Illustration of DICE score, adapted from Zou et al., (2004). E is the Expert contour P is the Participant contour. The DICE score is defined as $2\frac{V_{E\cap P}}{V_E+V_P}$ where $V_{E\cap P}$ is the volume of the intersected contours, V_E is the volume of the expert contour, and V_P is the volume of the participant contour.

Contours were reviewed by the expert and tutors responsible of Radiation Therapy Quality Assurance (RTQA) for the RAIDs project. Quantitatively, they were classified based on DICE index values given by the FALCON EduCase^{TM} output (**Figure 3.5**). The DICE score is defined as $2\frac{V_{E\cap P}}{V_E+V_P}$ where $V_{E\cap P}$ is the volume of the intersected contours, V_E is the volume of the expert contour, and V_P is the volume of the participant contour. The DICE score has values between 0 (no overlap between contours) and 1 (perfect overlap between contours) as illustrated in **Figure 3.5**.

- References for TV (Dimopoulos et al., 2009; Petersen et al., 2007):
  - A: Optimal: > 0.81
  - B: Average: 0.65 – 0.81
  - C: Suboptimal: < 0.65
- References for OAR (Breunig et al., 2012):
  - A: Optimal: > 0.81
  - B: Suboptimal: $<= 0.81$

In MRI-based BT for cervical cancer, Dimopoulos et al., (2009) determined a range of 0.5-0.7 using the conformity index for TV, which converted to DICE is approximately 0.625-

0.81 (Dimopoulos et al., 2009; Fotina et al., 2012). For OAR, Breunig et al., (2012) obtained an average DICE of 0.61 for volumes < 8 cm^3 and of 0.91 for volumes > 8 cm^3, averaging at 0.76 (Breunig et al., 2012). To simplify the cutoffs, making the study easier to interpret, 0.65 and 0.81 were chosen for TV and 0.81 for OAR. Importantly, all statistical analyses performed in this study were independent of the thresholds which were only used to aid interpretation and to display the results.

To perform the objective qualitative intraobserver assessment, the FALCON EduCase^{TM} contour error distance tool revealed on axial slices where the participant contour was 3 mm larger or smaller than the expert contour, based on the scalar assessment in the transverse plane for high risk CTV (HR-CTV) by Petric et al., (2008), in eight directions (anterior, posterior, right, left, anterolateral right and left and posterolateral right and left) to identify the most prevalent areas of uncertainties (Petric et al., 2008).

Qualitative Classification:

Correct: Participant contour $\leq$ 3 mm smaller/larger than the expert contour in a given direction.

Incorrect: Participant contour $>$ 3 mm smaller/larger than the expert contour in a given direction without a probable clinical impact.

Very incorrect: Participant contour $>$ 3 mm smaller/larger than the expert contour in a certain direction which for that particular ROI will have a probable clinical impact (worse coverage of TV/ higher dose to OAR).

Considering that part of the outcome of e-learning courses depends not only on the quality of the course itself, but also on how the participant perceives and understands it, a final anonymous satisfaction questionnaire adapted from FALCON ESTRO ODW was administered to the clinicians (see **Section 8.2**).

3.5 Statistical analysis

3.5.1 Preprocessing of the data

All the DICE scores have been transformed using the logit function, $\text{logit}(x) = \frac{x}{1-x}$. This transformation allows the DICE scores to asymptotically follow a Gaussian distribution (Agresti, 2002; Brock, 2013; Zou et al., 2004) which is a prerequisite of the statistical models we used. As the slice number ranges vary from one ROI to another, the slice numbers have

been centred on the middle of the range and scale such that the all the ranges vary from -1 to +1 for each ROI. Therefore, when the slice effect is considered in a statistical model, this centering allows a comparison between the ROI in the middle of their range.

3.5.2 Quantitative analysis of the interobserver variability

Mixed linear model

A mixed linear model was used to assess interobserver variability and fixed effects on the DICE score. The model in **Equation 3.1** was used to analyse the EBRT treatment and the model in **Equation 3.2** was used for the BT treatment and will be referred as to **modelINTER.PART**. They are essentially the same models, except that for EBRT the effect on the imaging technique was not included as it is not relevant (all institutions use the same imaging technique).

$$\begin{aligned} \text{logit}(d_{ijkmno}) &= \mu + \alpha_i + \beta_j + \gamma_k + \\ &\quad \alpha\beta_{ij} + \\ &\quad \alpha\gamma_{ik} + \\ &\quad \rho_i s_o + \nu_i s_o^2 + \\ &\quad z_{mn} + \\ &\quad e_{ijkmno} \end{aligned} \tag{3.1}$$

$$\begin{aligned} \text{logit}(d_{ijklmno}) &= \mu + \alpha_i + \beta_j + \gamma_k + \lambda_l + \\ &\quad \alpha\beta_{ij} + \\ &\quad \alpha\gamma_{ik} + \\ &\quad \alpha\lambda_{il} + \\ &\quad \rho_i s_o + \nu_i s_o^2 + \\ &\quad z_{mn} + \\ &\quad e_{ijklmno} \end{aligned} \tag{3.2}$$

where:

μ is the Intercept of the model

α_i is the fixed effect of ROI type i

β_j is the fixed effect of the contouring period j, $j \in \{C1, C2, C3\}$

γ_k is the fixed effect of a participant's experience k, $k \in \{AI, AE, JR, SR\}$

λ_l is the fixed effect of the imaging technique l, $l \in \{MRI, OTHER\}$

$\alpha\beta_{ij}$ is the interaction the between fixed effects ROI i and contouring period j

$\alpha\gamma_{ik}$ is the interaction the between fixed effects ROI i and the experience k

$\alpha\lambda_{il}$ is the interaction the between fixed effects ROI i and the imaging technique l

s_o is the linear fixed effect of the slice number for the observation o

s_o^2 is the quadratic fixed effect of the slice number for the observation o

ρ_i is the interaction between the ROI and the linear fixed effect of the slice number

ν_i is the interaction between the ROI and the quadratic fixed effect of the slice number

z_{mn} is the random effect of the participant m for the ROI type n, $n \in \{OAR, TV\}$

o is the observation index

$e_{ijklmno}$ is the residual error of the model

We assume that the values z_{mn} of the participant random effect follow a Gaussian distribution with a mean equal to 0 and a variance equal to σ^2_{TV}, for $n = TV$, and equal to σ^2_{OAR}, for $n = OAR$. Therefore, we consider in the model that the interobserver variability is different between the TV and the OAR. The values $e_{ijklmno}$ follow a Gaussian distribution with a mean equal to 0 and a variance equal to σ^2_R. Fixed effects are noted in Greek characters (ROI type, contouring period, experience, imaging technique for BT, linear and quadratic effects of slice number) while the random effect is noted in Latin characters (participant effect for the ROI type). The purpose of the mixed model is to estimate both fixed and random effects at the same time. The random part of the model allows the decomposition of the variance between different components: the variance due to the interobserver variability and the residual variance (which is the unexplained variance by the model).

We used the *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2016) to estimate the parameters of both models.

Pairwise comparisons

As the model considers many different effects and the design of the study is unbalanced (we do not have the same number of observations for the different conditions), we used the least-squares means (in short lsmeans) to obtain the average of an effect of interest adjusted over all the other effects in the model. For example, if we want to obtain the means of the DICE scores by ROI for each contouring period, the means will be averaging using the estimated means over the other effects. For this purpose, we used the *lsmeans* package (Lenth, 2016).

For all pairwise comparisons we corrected the *p-values* for multiple testing using false discovery rate (FDR) (Benjamin and Hochberg, 1995).

Intra-class correlation

The repeatability is an important concept to assess the accuracy of measurements. It expresses the proportion of the total variation that is reproducible among repeated measurements of the same criteria between different observers (Nakagawa and Schielzeth, 2010). The repeatability is often called the intra-class correlation (ICC) and this term will be used from here on. The ICC varies between 0 (no agreement between observers) and 1 (perfect agreement between observers). The **modelINTER.PART** considers that the ICC is different for TV and OAR and will be noted $ICC.TV$ and $ICC.OAR$, respectively:

$$\begin{aligned} ICC.TV &= \frac{\sigma^2_{TV}}{\sigma^2_{TV} + \sigma^2_R} \\ ICC.OAR &= \frac{\sigma^2_{OAR}}{\sigma^2_{OAR} + \sigma^2_R} \end{aligned}$$

The confidence intervals of $ICC.TV$ and $ICC.OAR$ were estimated using bootstrap with 1000 random permutations.

3.5.3 Quantitative analysis of the intraobserver variability

Linear model to assess the difference of DICE score by participant

To assess intraobserver variability between the different contouring periods, a linear model was used. From the model in **Equation 3.3** referred as to **modelINTRA.PART**, we estimated the differences of the means of the DICE scores (in logit scale) by participant and

by ROI over all the slices for all the three pairwise comparisons of the contouring period j with respect to j', $j > j', j$ and $j' \in \{C1, C2, C3\}$ separately.

$$\text{logit}(d_{ijmo}) - \text{logit}(d_{ij'mo}) \quad = \quad \mu + \alpha_i + \zeta_m + \alpha\zeta_{im} + e_{imo} \tag{3.3}$$

where:

μ is the Intercept of the model

α_i is the fixed effect of ROI i

ζ_m is the fixed effect of the participant m

o is the observation index

e_{imo} is the residual error of the model

The values e_{ijmo} follow a Gaussian distribution with a mean equal to 0 and a variance equal to σ_R^2. We used the least-squares means to estimate the parameters of the model.

Linear model to assess the DICE score by participant

To asses the DICE score by participant, a linear model with the same effects as the **Equation 3.3** was used. From the model in **Equation 3.4** referred as to **modelSCORE.PART**, we estimated the means of the DICE scores (in logit scale) by participant and by ROI over all the slices for every contouring period j, $j \in \{C1, C2, C3\}$ separately.

$$\text{logit}(d_{ijmo}) \quad = \quad \mu + \alpha_i + \zeta_m + \alpha\zeta_{im} + e_{imo} \tag{3.4}$$

Fisher's exact test to assess if the number of participants who improved is significant

From the model in **Equation 3.3**, we have a *p-value* by participant and by ROI. Those *p-values* assess the significance of the differences of the means of the DICE scores (in logit scale) between two consecutive contouring periods. As we used bilateral intervals with 95% confidence, we consider the difference being significant if the *p-value* is lower than 5%. As we have at lot of *p-values*, we expected to have 2.5% (since the interval is bilateral) of the participants to improve by chance under the null hypothesis. We test if the proportion of

participants who improved is significantly greater than 2.5%. Note that we used the same approach for participants performing worse (since we expected to have significant *p-values* for some participants but being just false negatives). These two tests will be noted $FEtest_{better}$ and $FEtest_{worse}$.

Fisher's exact test to assess the relationship between the participants who improved with respect to other covariates

From the model in **Equation 3.3** and the *p-values* obtained, we defined by ROI and by participant two categories:

1. Better
2. Not better (*i.e.* same or worse)

Using a Fisher's exact test, we assessed whether these two categories where associated with the other following covariates: inst, exp, img or ROI.type. These tests will be noted $FEtest_{inst}$, $FEtest_{exp}$, $FEtest_{img}$ and $FEtest_{ROI.type}$ respectively.

3.5.4 Qualitative analysis of the intraobserver variability

In **Section 3.4**, three qualitative classifications of the contours have been defined (*correct*, *incorrect* and *very incorrect*). In this section, we defined statistical models to assess whether the participants improved between the different contouring periods by comparing the proportions of *correct* contours. The proportions of correct contours were calculated by ROI (p_{ijv}) or by ROI.type (p_{njv}) for each contouring period and each variable $v \in$ {alatl, alatr, ant, inf, l, platl, platr, post, r, sup}. To assess whether there was a difference between the proportions between the different contouring periods (for example, are the proportions $p_{gtv,C1,inf}$ and $p_{gtv,C2,inf}$ different?), we used the test of McNemar (McNemar, 1947). It is a chi-square based statistic used to compare proportions on paired data (since we want to assess whether the participant improved between the different contouring periods).

3.5.5 Statistical software and reproductibility

We used R software (R Core Team, 2016) for statistical data analysis and *ggplot2* for graphics (Wickham, 2009). An automated reproduction of this analysis may be performed by using the scripts and data included in the supplementary information provided in the article Rivin del Campo et al., (2017).

4 Results

Contents

4.1 Descriptive statistics

Seventeen participating Rational molecular Assessments and Innovative Drug Selection (RAIDs) centres answered the preworkshop questionnaire assessing their everyday practice (**Table 4.1**). Contours were submitted by participants from 14 of 22 RAIDs centres (**Table 4.1**).

Forty-six participants were enrolled of which nine submitted delineations for all contouring periods for external beam radiotherapy (EBRT) and brachytherapy (BT) (**Table 4.2**). A description of the participant population which submitted contours by their level of experience is presented in **Table 4.3**.

The histograms of the DICE values (in logit scale) for EBRT and BT treatments are presented in **Figure 4.1**.

Table 4.1: Preworkshop questionnaire reflecting daily practice in each centre (Rivin del Campo et al., 2017).

Centre	Centre Type	# Pat. RCT/Yr.	Type C. Ca. Pat. Treated	BT Applicator	Do Interstitial BT	Dose Rate	BT Imaging Type	Plan To Start 3D IGBT?	BT Prescription	Dose TV	# Fract.	Constraints
Centre 1•	Academic	<50	All pat. S.I–IVa, M0	Ovoids; Interstitial N.	YES	PDR	CT; MRI; US	NA	HR-CTV	80–84 Gy	2	Bladder Rectum Sigmoid
Centre 2•	Public	50–100	All pat. S.I–IVa, M0	Ovoids	NO	HDR	CBCT	In 3y	PointA	65–74 Gy	4	Bladder Rectum
Centre 3•	Academic; Public	>100	All pat. S.I–IVa, M0	Tandem/ring; Ovoids; Tandem/cylinder	NO	HDR	X-ray	NO	PointA	65–74 Gy	4	Bladder Rectum
Centre 4•	Academic	<50	All pat. S.I–IVa, M0	Ovoids	NO	PDR	CT; MRI	NA	PointA; HR-CTV	65–74 Gy	2	Bladder Rectum Sigmoid
Centre 5•	Academic; Public	>100	All pat. S.I–IVa, M0	Ovoids	NO	HDR	CT; MRI	NA	HR-CTV	75–79 Gy	2	Bladder Rectum Sigmoid
Centre 6•	Academic	<50	Pos. Pelv./PA LN	Ovoids	YES	PDR	CT; MRI	NA	HR-CTV	80–84 Gy	2	Bladder Rectum Sigmoid
Centre 7•	Private	50–100	All pat. S.I–IVa, M0	Ovoids; Interstitial N.	YES	PDR; LDR	CT; MRI; US	NA	HR-CTV	<65 Gy; 65–74 Gy	-	Bladder Rectum Sigmoid
Centre 8•	Academic	<50	Pos. Pelv./PA LN; Neg.LN S. > IIB	Ovoids; Mould; Interstitial N.	YES	PDR	CT; MRI	NA	HR-CTV	<65 Gy	2	Bladder Rectum Sigmoid
Centre 9•	Public	<50	Pos. Pelv./PA LN; Neg.LN S. > IIB	Tandem/ring; Mould	NO	PDR	CT; US	NA	HR-CTV	80–84 Gy	2	Bladder Rectum Sigmoid
Centre 10•	Academic; Public	<50	Pos. Pelv./PA LN; Neg.LN S. > IIB	Tandem/ring; Ovoids	NO	PDR	CT; US	NA	PointA; HR-CTV; IR-CTV	80–84 Gy	2	Bladder Rectum Sigmoid
Centre 11	Public	<50	Pos. PA LN; Neg.LN S. > IIB	Mould	NO	LDR	CT; MRI	NA	IR-CTV	65–74 Gy	2	Bladder Rectum
Centre 12•	Public	<50	All pat. S.I–IVa, M0	Ovoids	NO	PDR	CT	NA	HR-CTV	65–74 Gy	2	Bladder Rectum Sigmoid
Centre 13•	Academic	50–100	Pos. Pelv./PA LN; Neg.LN S. > IIB	Ovoids	NO	PDR	CT; MRI	NA	HR-CTV; IR-CTV	<65 Gy; 80–84 Gy	2	Bladder Rectum Sigmoid Vagina
Centre 14•	Academic	>100	All pat. S.I–IVa, M0	Ovoids; Mould; Interstitial N.	YES	PDR	MRI	NA	HR-CTV	>85 Gy	2	Bladder Rectum Sigmoid
Centre 15	Public	<50	Pos. Pelv./PA LN	Ovoids	YES	HDR	MRI	NA	HR-CTV	>85 Gy	3	Bladder Rectum Sigmoid
Centre 16	Academic	50–100	All pat. S.I–IVa, M0	Ovoids	YES	PDR	MRI	NA	HR-CTV	75–79 Gy; 80–84 Gy	2	Bladder Rectum Sigmoid
Centre 17•	Academic	>100	All pat. S.I–IVa, M0	Ovoids; Tandem/cylinder	NO	HDR	CT	NA	PointA; HR-CTV	80–84 Gy	2; 3	Bladder Rectum

The seventeen centres listed below answered this questionnaire, of which the 14 marked with an asterisk participated in the ODW (the order of the centres does not correspond with the order in the anonymous table).
French centres: Institut Curie•, Gustave Roussy•, Centre Georges François Leclerc•, Centre Léon Bérard•, Centre Alexis Vautrin•, CHU Anne de Bretagne•, Centre Jean Perrin•, Institut Bergonié•, Institut de Cancérologie de l'Ouest•; Centre Oscar Lambret•, Tenon Hospital, Institut Claudius Regaud.
The Netherlands: Academic Medical Centre (AMC)•, Antoni van Leewenhoek Hospital (NKI).
Moldova: Institute of Oncology (IOM)•.
Romania: Emergency County Hospital Oradea•.
Serbia: Institut za onkologiju Vojvodine (IOV)•.
Abbreviations: RCT: Radiochemotherapy; yr.: year; C.Ca.: Cervical cancer; Pat.: patients; BT: Brachytherapy; IGBT: Image guided brachytherapy; TV: Target Volume; Fract.: Fractions; S.: Stage; Pos.: Positive; Pelv.: Pelvic; PA: Para-aortic; LN: Lymph Nodes; N.: Needles; LDR: Low Dose Rate; PDR: Pulsed Dose Rate; HDR: High Dose Rate; CT: Computed Tomography scan; MRI: Magnetic Resonance Imaging; CBCT: Cone Beam CT scan; US: Ultrasound; NA: Not applicable; HR-CTV: High Risk CTV; IR-CTV: Intermediate Risk CTV; Gy: Gray.

	EBRT			BT		
	C1	**C2**	**C3**	**C1**	**C2**	**C3**
Submission of ≥ 1 contour	28	22	13	30	21	13
Submission of all contours (OAR and TV)	14	11	5	24	15	6
Submission of only TV	19	15	7	24	15	9
Total number of participants (n)			46			46

Table 4.2: Participants enrolled in the online delineation workshops which submitted contours for each contouring period. Abbreviations: EBRT: External beam radiotherapy; BT: Brachytherapy; OAR: Organs at risk; TV: Target volumes.

PARTICIPANT POPULATION WHICH SUBMITTED CONTOURS	
Experienced specialists (>5 years of post-residency experience)	13
Less experienced specialists (≤5 years of post-residency experience)	8
Senior residents (>2 years of experience)	7
Junior residents (≤2 years of experience)	4

Table 4.3: Description of the participant population which submitted contours, by level of experience.

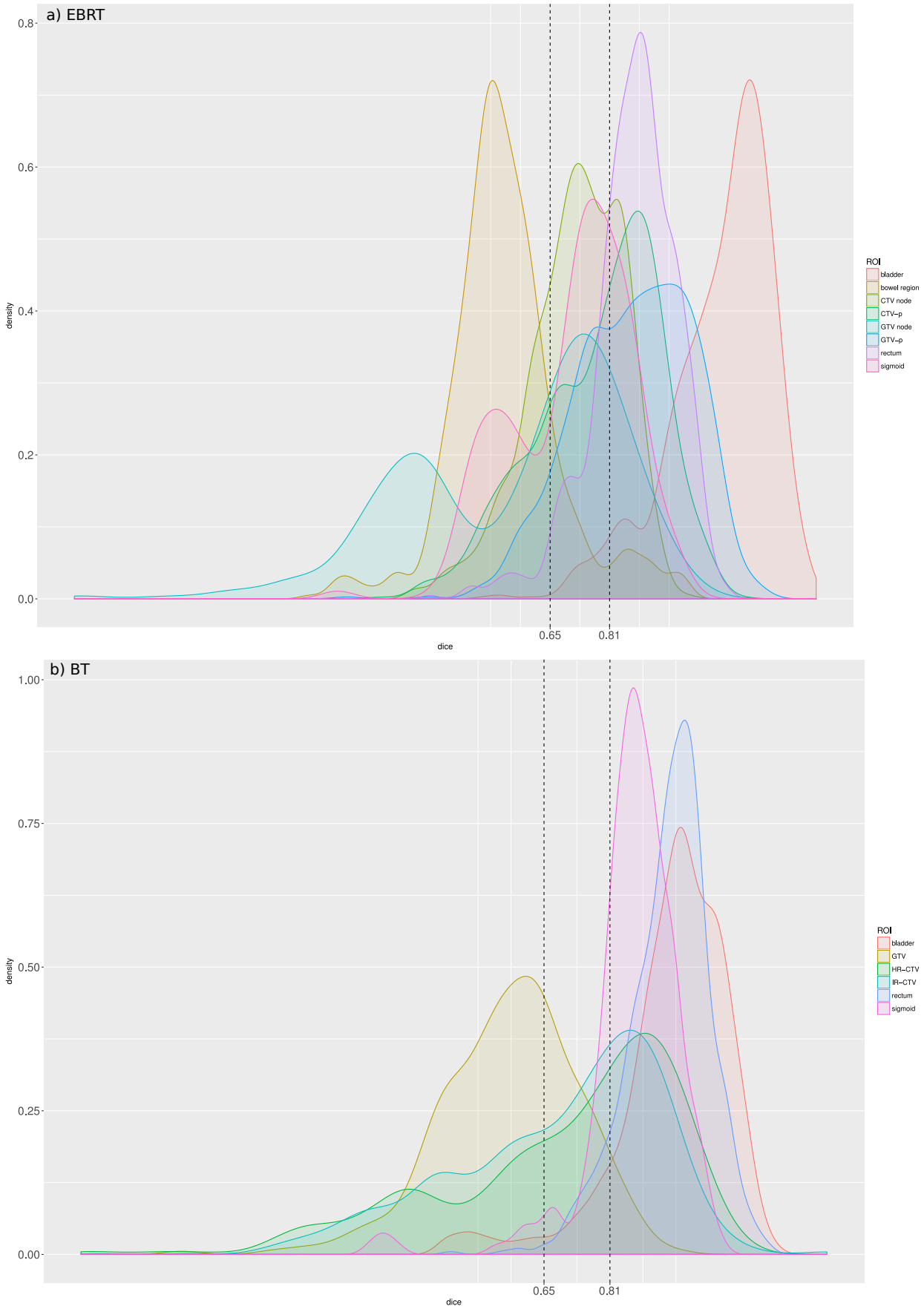


Figure 4.1: Distribution of DICE scores (in logit scale) for EBRT and BT. The thresholds 0.65 and 0.81 (used to define the optimality of contours) are indicated in the logit scale.

4.2 Noteworthy statistical specifications and general summary of results

4.2.1 Analysis of the participants who left the study

We have defined the variable *has.left* to identify the participants who participated to final contouring (C3) (*has.left* = *NO*) and participants who left the study, whether it be after baseline contouring (C1) or guideline contouring (C2) (*has.left* = *YES*). Using the same linear mixed model (without the contouring period effect), we assessed the effect of the variable *has.left* on the DICE score at C1 contouring attempt, to evaluate whether the baseline DICE score had an influence on participants adherence to the online delineation workshop (ODW). Using Fisher's exact test, we also assessed the association between this variable and the centre or the experience. Importantly, for both EBRT and BT, the **Table 4.4** shows that the variable *has.left* has no significant effect on the DICE score at C1, and no association with the centre or experience (all the *p-values* are > 0.05).

Hypothesis H_0	**Statistical test**	**EBRT**	**BT**
Has the variable *has.left* an effect on the DICE score at C1?	Fisher's test of type III with Satterthwaite approximation for degrees of freedom	0.8294	0.5892
Is there an association between the variable *has.left* and the institution?	Fisher's exact test	0.10	0.11
Is there an association between the variable *has.left* and the position?	Fisher's exact test	0.70	0.72

Table 4.4: *P-values* for the different hypotheses tested to assessed the effect of the variable *has.left* for both EBRT and BT. The answer to the Hypothesis tested is *Yes* when the *p-value* is < 0.05.

4.2.2 Cutoff points for the DICE scores

A verification of the adequacy of the cutoff points used in this study was performed by examining the distribution of the pooled data for both EBRT and BT treatments over all contouring periods. In the case of the organs at risk (OAR), the first quartile is 0.8, and for target volume (TV) the first quartile is 0.6 and the third quartile is 0.86. These quartiles are for the most part consistent with the employed cutoff points (0.65 and 0.81).

4.2.3 Summary of the results

A summary of the main results are presented in **Table 4.5**.

	EBRT			BT		
INTEROBSERVER QUANTITATIVE						
	C2 vs. C1	**C3 vs. C1**	**C3 vs. C2**	**C2 vs. C1**	**C3 vs. C1**	**C3 vs. C2**
Comparisons between contouring periods by ROI•	↑Bowel		↓Bowel	↑Bladder		↓Bladder
	↑CTV node	↑CTV node	↑CTV node	↑GTV		
	↑CTV-p		↓CTV-p	↑HR-CTV	↑HR-CTV	
	↑GTV node	↓GTV node	↓GTV node	↑IR-CTV	↑IR-CTV	
		↑GTV-p				
Comparisons between experience by ROI•	Sigmoid: Exp. Spec. > Junior Res. Less exp. Spec. > Junior Res. Less exp. Spec. > Senior Res. GTV node: Exp. Spec. < Less exp. Spec. Exp. Spec. < Senior Res. Less exp. Spec. > Junior Res. Senior Res. > Junior Res. GTV-p: Exp. Spec. < Junior Res. Junior Res. > Senior Res.			Sigmoid: Exp. Spec. > Senior Res.		
Comparison between imaging techniques•				HR-CTV: C. MRI-IGBT > Others		
INTRAOBSERVER QUANTITATIVE						
	C2 vs. C1	**C3 vs. C1**	**C3 vs. C2**	**C2 vs. C1**	**C3 vs. C1**	**C3 vs. C2**
Do participants improve between contouring periods?	Yes	Yes	Yes	Yes		
Is the improvement associated to ROI.type (TV/OAR)?	Yes			Yes		
Is the improvement associated to institution?	Yes					
INTRAOBSERVER QUALITATIVE (comparison of the % of correct contours between contouring periods)						
TV Posterolat. right	↑			↑	↓	↓
TV Anterolat. right		↓	↓			
TV Posterior			↓			
TV Posterolat. left			↓			
TV Right				↑		
OAR Posterolat. right					↑	

Abbreviations: EBRT: External beam radiotherapy; BT: Brachytherapy; C1: baseline contouring; C2: guideline contouring; C3: final contouring; exp.: experience; Spec.: Specialist; Res.: Resident; C. MRI-IGBT: centres using MRI-image guided BT; Posterolat. : Posterolateral; Anterolat. : Anterolateral.
↑: Improvement; ↓: Decrease.
• $p < 0.05$ after correction for multiple testing FDR.

Table 4.5: Results for both the interobserver and intraobserver quantitative / qualitative analyses. All of the reported results were statistically significant ($p\text{-}values < 0.05$) (Rivin del Campo et al., 2017).

4.3 Quantitative interobserver variability

This section presents the results of **modelINTER.PART** from **Equation 3.1** and **Equation 3.2** for EBRT and BT, respectively.

4.3.1 Evaluation of the different effects and their interactions on the DICE score

All of the interactions analysed were highly significant (*p-values* < 0.001) and all effects had a repercussion on DICE scores (**Table 4.6** and **Table 4.7** give the Type III sum of squares and associated *p-values* for the all the effects included in the models).

	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr(>F)
ROI	3311.58	473.08	7.00	390.71	1057.82	0.0000
c	10.43	5.22	2.00	6100.97	11.66	0.0000
exp	0.87	0.29	3.00	25.68	0.65	0.5910
slice	0.05	0.05	1.00	7464.78	0.12	0.7309
slice2	419.67	419.67	1.00	7464.26	938.39	0.0000
ROI:c	91.67	6.55	14.00	7211.05	14.64	0.0000
ROI:slice	349.89	49.98	7.00	7452.21	111.77	0.0000
ROI:slice2	170.73	24.39	7.00	7459.58	54.54	0.0000
ROI:exp	213.20	10.15	21.00	291.47	22.70	0.0000

Table 4.6: Analysis of Variance Table (on the fixed effects) of type III with Satterthwaite approximation for degrees of freedom for EBRT (Rivin del Campo et al., 2017). The model is computed on logit scale and corresponds to **Equation 3.1**. Abbreviations - Sum Sq: sum of squares of the effect; Mean Sq: mean squares of the effect (Mean Sq = Sum Sq / NumDF); NumDF: Degrees of freedom of the numerator of the Fisher's test statistics; DenDF: Degrees of freedom of the denominator of the Fisher's test statistics: F.value: value of the Fisher's test; Pr(>F): *p-value* of the Fisher's test.

	Sum Sq	Mean Sq	NumDF	DenDF	F.value	Pr(>F)
ROI	222.99	44.60	5.00	211.20	106.26	0.0000
c	11.62	5.81	2.00	3360.51	13.84	0.0000
exp	1.06	0.35	3.00	26.36	0.84	0.4817
img	1.58	1.58	1.00	26.31	3.77	0.0630
slice	45.39	45.39	1.00	3489.11	108.14	0.0000
slice2	436.40	436.40	1.00	3494.02	1039.77	0.0000
ROI:c	11.50	1.15	10.00	3467.90	2.74	0.0023
ROI:img	10.99	2.20	5.00	151.72	5.24	0.0002
ROI:slice	292.26	58.45	5.00	3487.88	139.27	0.0000
ROI:slice2	155.72	31.14	5.00	3487.63	74.20	0.0000
ROI:exp	25.99	1.73	15.00	152.42	4.13	0.0000

Table 4.7: Analysis of Variance Table (on the fixed effects) of type III with Satterthwaite approximation for degrees of freedom for BT. The model is computed on logit scale and corresponds to **Equation 3.2**. Abbreviations - Sum Sq: sum of squares of the effect; Mean Sq: mean squares of the effect (Mean Sq = Sum Sq / NumDF); NumDF: Degrees of freedom of the numerator of the Fisher's test statistics; DenDF: Degrees of freedom of the denominator of the Fisher's test statistics: F.value: value of the Fisher's test; Pr(>F): *p-value* of the Fisher's test.

The models from **Equation 3.1** and **Equation 3.2** clearly represent the quadratic relationship between DICE score and slice number, capturing efficiently the parabolic effect of the slice number on the DICE scores (**Figure 4.2**) for both EBRT and BT.

Figure 4.2: Means of the DICE scores (1 being perfect concordance between the participant and the expert; 0 being no concordance) along the three contouring periods, of all participants, by ROI according to the slice number. The lines correspond to the mean of the DICE score as predicted by the mixed models (from **Equation 3.1** and **Equation 3.2** for EBRT and BT respectively) while the dots are the means estimated from the raw data without any statistical model (Rivin del Campo et al., 2017).

4.3.2 Pairwise comparisons

The pairwise comparisons and corresponding plots are given in:

- **Figure 4.3** and **Table 4.8** for ROI and contouring period for EBRT
- **Figure 4.4** and **Table 4.9** for ROI and contouring period for BT
- **Figure 4.5** and **Table 4.10** for ROI and experience for EBRT
- **Figure 4.6** and **Table 4.11** for ROI and experience for BT
- **Figure 4.7** and **Table 4.12** for ROI and imaging technique for BT

Comparison between contouring period for each ROI

The pairwise comparisons for both EBRT and BT treatments between contouring periods by ROI are reported in **Table 4.5** and detailed in **Table 4.8** and **Table 4.9**. For EBRT and BT in C2 vs. C1 there was significant improvement, for the most part for TV, with no significant decrease. When considering C3 vs. C1 in EBRT and BT there was also a significant increase for certain TV with only a significant decrease for radiologically pathological lymph nodes to boost (GTV-node). However, in C3 vs. C2 for both treatments there was a significant decrease for two TV in EBRT and two OAR (no decrease for TV in BT), with a significant increase for nodal elective volume (CTV-node).

comparison	estimate	*p-value*	
bladder,C1 - bladder,C2	-0.01	0.85	
bladder,C1 - bladder,C3	-0.04	0.51	
bowel region,C1 - bowel region,C2	-0.13	0.04	*
bowel region,C1 - bowel region,C3	0.10	0.22	
CTV node,C1 - CTV node,C2	-0.11	0.00	**
CTV node,C1 - CTV node,C3	-0.44	0.00	***
CTV-p,C1 - CTV-p,C2	-0.19	0.00	***
CTV-p,C1 - CTV-p,C3	0.05	0.30	
GTV node,C1 - GTV node,C2	-0.19	0.00	***
GTV node,C1 - GTV node,C3	0.42	0.00	***
GTV-p,C1 - GTV-p,C2	-0.05	0.36	
GTV-p,C1 - GTV-p,C3	-0.15	0.03	*
rectum,C1 - rectum,C2	-0.05	0.39	
rectum,C1 - rectum,C3	-0.10	0.24	
sigmoid,C1 - sigmoid,C2	-0.04	0.67	
sigmoid,C1 - sigmoid,C3	-0.11	0.35	
bladder,C2 - bladder,C3	-0.03	0.60	
bowel region,C2 - bowel region,C3	0.23	0.01	**
CTV node,C2 - CTV node,C3	-0.32	0.00	***
CTV-p,C2 - CTV-p,C3	0.25	0.00	***
GTV node,C2 - GTV node,C3	0.61	0.00	***
GTV-p,C2 - GTV-p,C3	-0.10	0.16	
rectum,C2 - rectum,C3	-0.05	0.58	
sigmoid,C2 - sigmoid,C3	-0.07	0.54	

Table 4.8: Pairwise comparisons between contouring period for each ROI for EBRT from model **Equation 3.1**. The estimates correspond to the difference between the two conditions in the logit scale. The *p-values* have been corrected for multiple testing using FDR. Significant *p-values* are indicated by *.

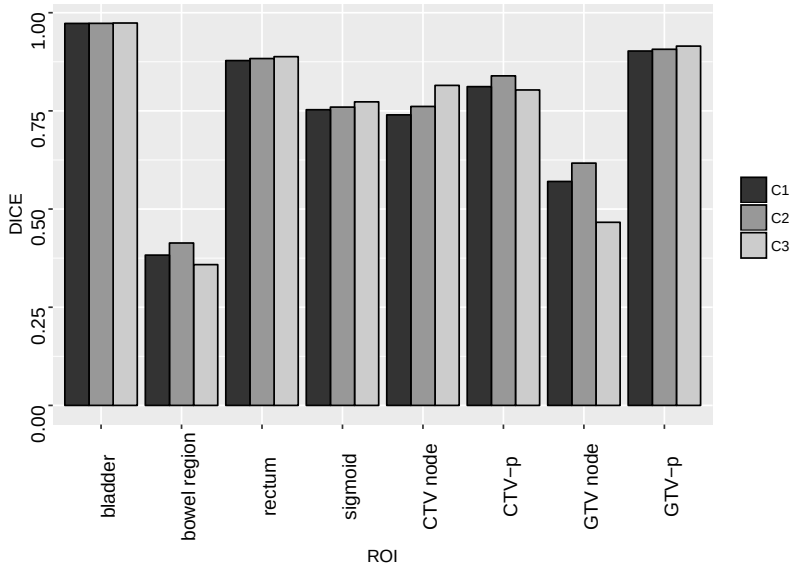


Figure 4.3: Lsmeans estimates by ROI and contouring period for EBRT. *P-values* for all pairwise comparisons are listed in **Table 4.8**.

comparison	estimate	*p-value*	
bladder,C1 - bladder,C2	-0.15	0.01	*
bladder,C1 - bladder,C3	0.12	0.12	
GTV,C1 - GTV,C2	-0.30	0.00	***
GTV,C1 - GTV,C3	-0.09	0.42	
HR-CTV,C1 - HR-CTV,C2	-0.24	0.00	***
HR-CTV,C1 - HR-CTV,C3	-0.16	0.02	*
IR-CTV,C1 - IR-CTV,C2	-0.19	0.00	***
IR-CTV,C1 - IR-CTV,C3	-0.16	0.01	*
rectum,C1 - rectum,C2	0.04	0.51	
rectum,C1 - rectum,C3	0.07	0.40	
sigmoid,C1 - sigmoid,C2	-0.07	0.47	
sigmoid,C1 - sigmoid,C3	-0.03	0.81	
bladder,C2 - bladder,C3	0.27	0.00	***
GTV,C2 - GTV,C3	0.21	0.07	
HR-CTV,C2 - HR-CTV,C3	0.08	0.31	
IR-CTV,C2 - IR-CTV,C3	0.03	0.65	
rectum,C2 - rectum,C3	0.03	0.72	
sigmoid,C2 - sigmoid,C3	0.04	0.74	

Table 4.9: Pairwise comparisons between contouring period for each ROI for BT from model **Equation 3.2**. The estimates correspond to the difference between the two conditions in the logit scale. The *p-values* have been corrected for multiple testing using FDR. Significant *p-values* are indicated by *.

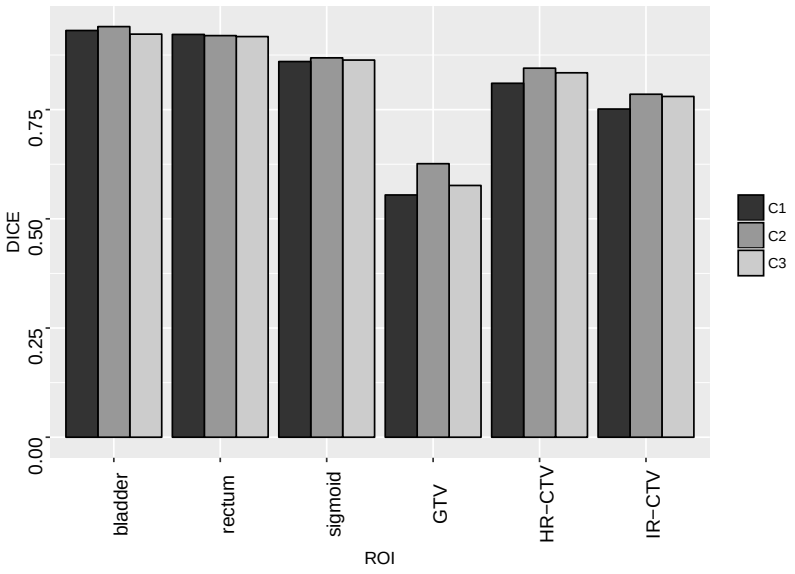


Figure 4.4: Lsmeans estimates by ROI and contouring period for BT. *P-values* for all pairwise comparisons are listed in **Table 4.9**.

Comparison between experience for each ROI

The analysis of the EBRT treatment (**Table 4.10**) regarding the experience effect showed that experienced specialists performed significantly better than junior residents for sigmoid and significantly worse than less experienced specialists and senior residents for GTV-node and than junior residents for cervix, parametria and vaginal gross disease (GTV-p). Less experienced specialists had performed significantly better than junior residents for GTV-node and sigmoid, and than senior residents for sigmoid. In the case of senior and junior residents, there were only significant differences between them for GTV-node and GTV-p. For BT the only significant difference was that experienced specialists had better results than senior residents for sigmoid (**Table 4.5** and **Table 4.11**).

comparison	estimate	*p-value*	
bladder,AE - bladder,AI	0.11	0.23	
bladder,AE - bladder,JR	0.01	0.93	
bladder,AE - bladder,SR	0.12	0.21	
bowel region,AE - bowel region,AI	-0.00	1.00	
bowel region,AE - bowel region,JR	-0.22	0.08	
bowel region,AE - bowel region,SR	-0.03	0.83	
CTV node,AE - CTV node,AI	-0.06	0.77	
CTV node,AE - CTV node,JR	0.16	0.52	
CTV node,AE - CTV node,SR	0.12	0.57	
CTV-p,AE - CTV-p,AI	0.07	0.75	
CTV-p,AE - CTV-p,JR	-0.08	0.78	
CTV-p,AE - CTV-p,SR	-0.33	0.11	
GTV node,AE - GTV node,AI	-0.61	0.00	**
GTV node,AE - GTV node,JR	0.41	0.11	
GTV node,AE - GTV node,SR	-0.96	0.00	***
GTV-p,AE - GTV-p,AI	-0.31	0.13	
GTV-p,AE - GTV-p,JR	-0.74	0.00	**
GTV-p,AE - GTV-p,SR	-0.14	0.53	
rectum,AE - rectum,AI	0.02	0.90	
rectum,AE - rectum,JR	0.07	0.58	
rectum,AE - rectum,SR	0.01	0.94	
sigmoid,AE - sigmoid,AI	-0.10	0.49	
sigmoid,AE - sigmoid,JR	0.39	0.01	*
sigmoid,AE - sigmoid,SR	0.19	0.14	
bladder,AI - bladder,JR	-0.10	0.40	
bladder,AI - bladder,SR	0.01	0.96	
bowel region,AI - bowel region,JR	-0.22	0.16	
bowel region,AI - bowel region,SR	-0.03	0.88	
CTV node,AI - CTV node,JR	0.23	0.40	
CTV node,AI - CTV node,SR	0.18	0.42	
CTV-p,AI - CTV-p,JR	-0.15	0.60	
CTV-p,AI - CTV-p,SR	-0.40	0.08	
GTV node,AI - GTV node,JR	1.02	0.00	***
GTV node,AI - GTV node,SR	-0.35	0.14	
GTV-p,AI - GTV-p,JR	-0.42	0.12	
GTV-p,AI - GTV-p,SR	0.18	0.45	
rectum,AI - rectum,JR	0.06	0.68	
rectum,AI - rectum,SR	-0.01	0.96	
sigmoid,AI - sigmoid,JR	0.49	0.00	**
sigmoid,AI - sigmoid,SR	0.29	0.04	*
bladder,JR - bladder,SR	0.11	0.38	
bowel region,JR - bowel region,SR	0.19	0.16	
CTV node,JR - CTV node,SR	-0.04	0.88	
CTV-p,JR - CTV-p,SR	-0.25	0.35	
GTV node,JR - GTV node,SR	-1.36	0.00	***
GTV-p,JR - GTV-p,SR	0.60	0.03	*
rectum,JR - rectum,SR	-0.06	0.65	
sigmoid,JR - sigmoid,SR	-0.20	0.23	

Table 4.10: Pairwise comparisons between experience for each ROI for EBRT from model **Equation 3.1**. The estimates correspond to the difference between the two conditions in the logit scale. The *p-values* have been corrected for multiple testing using FDR. Significant *p-values* are indicated by *.

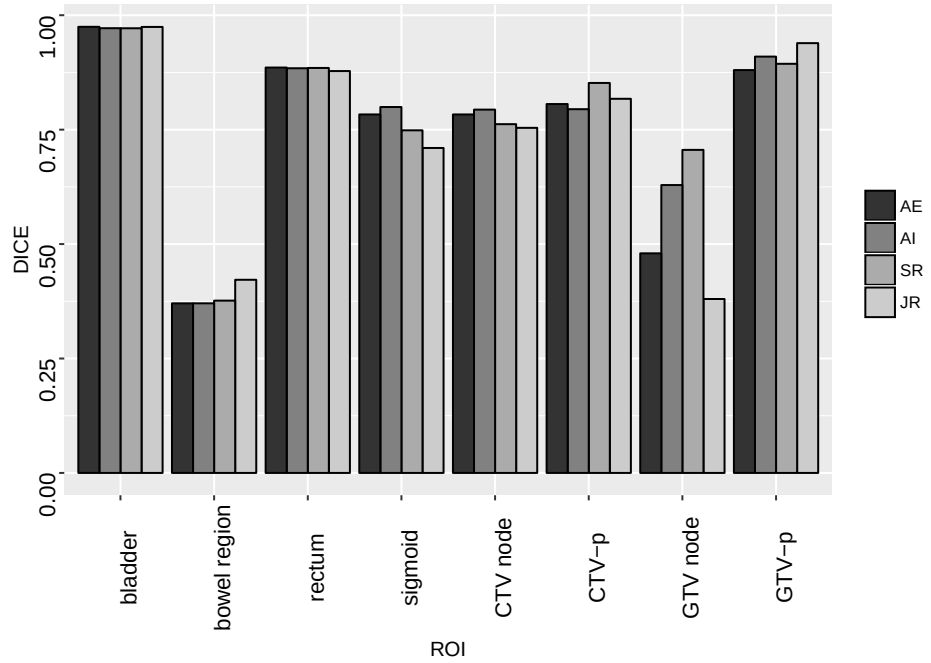


Figure 4.5: Lsmeans estimates by ROI and experience for EBRT. *P-values* for all pairwise comparisons are listed in **Table 4.10**.

comparison	estimate	*p-value*	
bladder,AE - bladder,AI	0.05	0.76	
bladder,AE - bladder,JR	0.02	0.90	
bladder,AE - bladder,SR	-0.01	0.95	
GTV,AE - GTV,AI	0.20	0.67	
GTV,AE - GTV,JR	-0.13	0.83	
GTV,AE - GTV,SR	0.19	0.69	
HR-CTV,AE - HR-CTV,AI	0.70	0.10	
HR-CTV,AE - HR-CTV,JR	0.58	0.26	
HR-CTV,AE - HR-CTV,SR	0.13	0.79	
IR-CTV,AE - IR-CTV,AI	0.64	0.13	
IR-CTV,AE - IR-CTV,JR	0.52	0.31	
IR-CTV,AE - IR-CTV,SR	0.09	0.85	
rectum,AE - rectum,AI	0.24	0.06	
rectum,AE - rectum,JR	0.16	0.31	
rectum,AE - rectum,SR	0.20	0.16	
sigmoid,AE - sigmoid,AI	0.20	0.21	
sigmoid,AE - sigmoid,JR	0.20	0.32	
sigmoid,AE - sigmoid,SR	0.39	0.02	*
bladder,AI - bladder,JR	-0.02	0.91	
bladder,AI - bladder,SR	-0.05	0.73	
GTV,AI - GTV,JR	-0.32	0.57	
GTV,AI - GTV,SR	-0.01	0.99	
HR-CTV,AI - HR-CTV,JR	-0.12	0.84	
HR-CTV,AI - HR-CTV,SR	-0.57	0.21	
IR-CTV,AI - IR-CTV,JR	-0.12	0.84	
IR-CTV,AI - IR-CTV,SR	-0.55	0.23	
rectum,AI - rectum,JR	-0.08	0.67	
rectum,AI - rectum,SR	-0.04	0.79	
sigmoid,AI - sigmoid,JR	0.00	0.99	
sigmoid,AI - sigmoid,SR	0.19	0.29	
bladder,JR - bladder,SR	-0.03	0.87	
GTV,JR - GTV,SR	0.32	0.59	
HR-CTV,JR - HR-CTV,SR	-0.45	0.42	
IR-CTV,JR - IR-CTV,SR	-0.43	0.45	
rectum,JR - rectum,SR	0.04	0.85	
sigmoid,JR - sigmoid,SR	0.19	0.40	

Table 4.11: Pairwise comparisons between experience for each ROI for BT from model **Equation 3.2**. The estimates correspond to the difference between the two conditions in the logit scale. The *p-values* have been corrected for multiple testing using FDR. Significant *p-values* are indicated by *.

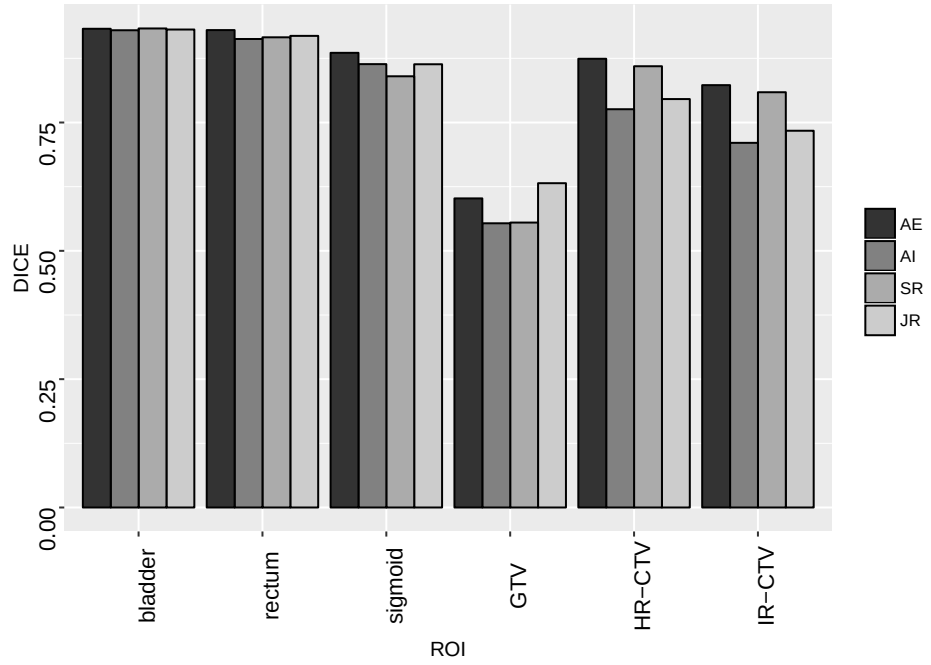


Figure 4.6: Lsmeans estimates by ROI and experience for BT. *P-values* for all pairwise comparisons are listed in **Table 4.11**.

Effect of the MRI-imaging technique for the BT treatment

Regarding the imaging technique, the **INTER.PART** model from **Equation 3.2** was used to analyse interobserver variability between centres that used magnetic resonance imaging (MRI) based image guided brachytherapy (IGBT) and those which did not. Centres using MRI based IGBT did significantly better than those which used other techniques (computed tomography (CT), X-Ray, ultrasound (US)) for high risk CTV (HR-CTV) (**Table 4.12**, **Figure 4.7**).

comparison	estimate	*p-value*	
bladder,MRI - bladder,OTHER	0.00	0.97	
GTV,MRI - GTV,OTHER	0.41	0.21	
HR-CTV,MRI - HR-CTV,OTHER	0.78	0.02	*
IR-CTV,MRI - IR-CTV,OTHER	0.53	0.10	
rectum,MRI - rectum,OTHER	0.16	0.12	
sigmoid,MRI - sigmoid,OTHER	0.18	0.15	

Table 4.12: Pairwise comparisons between imaging technique for each ROI for BT from model **Equation 3.2**. The estimates correspond to the difference between the two conditions in the logit scale. The *p-values* have been corrected for multiple testing using FDR. Significant *p-values* are indicated by *.

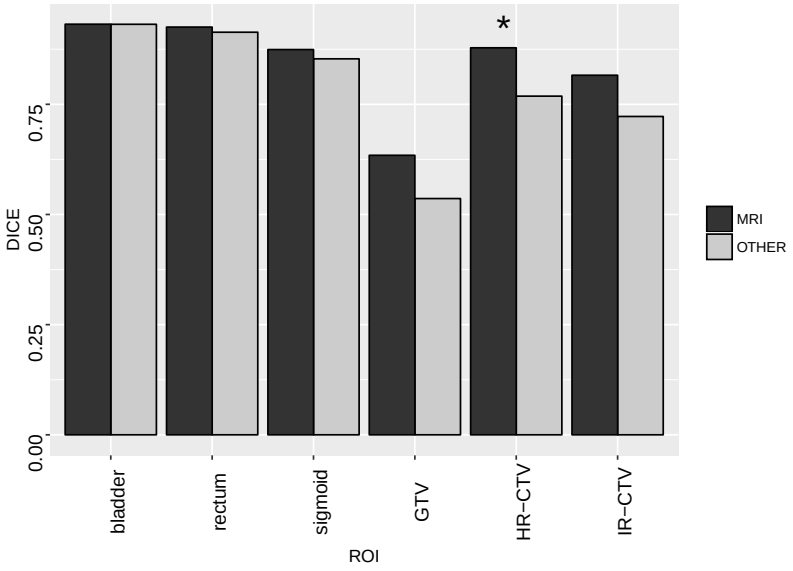


Figure 4.7: Lsmeans estimates by ROI and imaging technique. Significant difference is indicated by *. *P-values* for all pairwise comparisons are listed in **Table 4.12**.

Intra-class correlation

The intra-class correlation (ICC) for interobserver variability was excellent for OAR in BT (0.92; 95% confidence interval: 0.86-0.96), OAR in EBRT (0.96; 95% confidence interval: 0.93-0.98) and TV in EBRT (0.78 - 95% confidence interval: 0.68-0.88) while it was fair for

TV in BT (0.51; 95% confidence interval: 0.39-0.68). The low ICC for TV in BT highlights the difficulty of participants to agree on contours, whether they usually contour on MRI or not (the imaging technique was taken into account in **modelINTER.PART**).

4.4 Quantitative intraobserver variability

This section presents the results of **Equation 3.3** referred as to **modelINTRA.PART** and **Equation 3.4** referred as to **modelSCORE.PART**.

4.4.1 Difference of the DICE scores between C2 vs. C1

Figure 4.8 and **Figure 4.9** show the difference of DICE scores by participant between C2 vs. C1 by region of interest (ROI) for EBRT and BT, respectively (only the plots for the comparison between C2 vs. C1 are reported in the present manuscript to facilitate the lecture of this thesis, all the other plots for C3 vs. C2 and C3 vs. C1 are available in the supplementary materials in Rivin del Campo et al., 2017). The values plotted correspond to the estimates obtained by **modelINTRA.PART** from **Equation 3.3**. The participants improve when the lsmean value is positive and the confidence interval does not overlap 0. The **Table 4.13** provides the different Fisher's exact tests to assess the significance of the improvement and its association with other effects.

The **Figure 4.10** represents the scatterplot of the DICE score between C2 vs. C1.

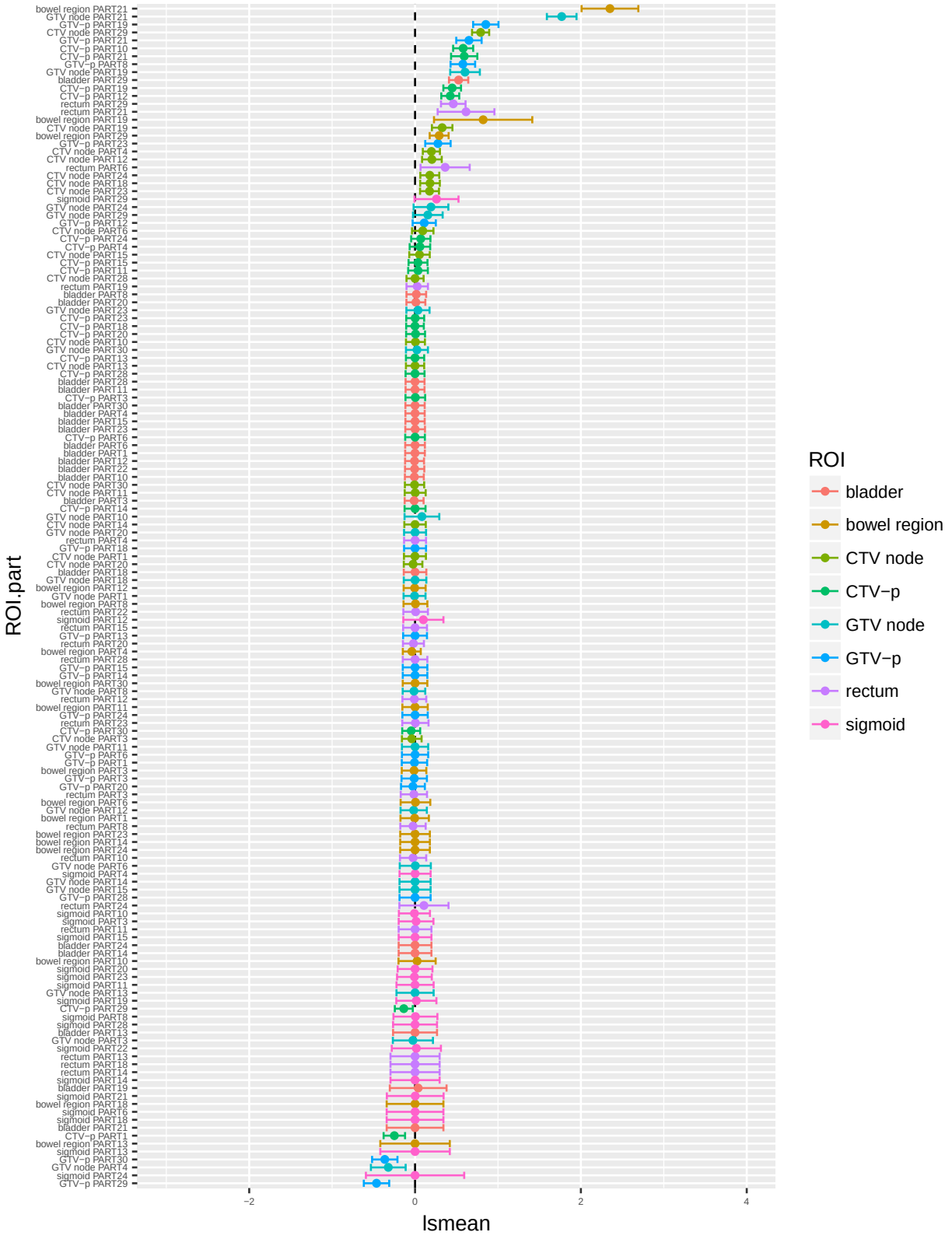


Figure 4.8: Difference of the DICE scores by participant between C2 vs. C1 by ROI. The values are the lsmeans estimates with 95% confidence interval (logit scale) obtained from **modelINTRA.PART** for EBRT. Participants improve when the lsmean value is positive and the confidence interval does not overlap 0.

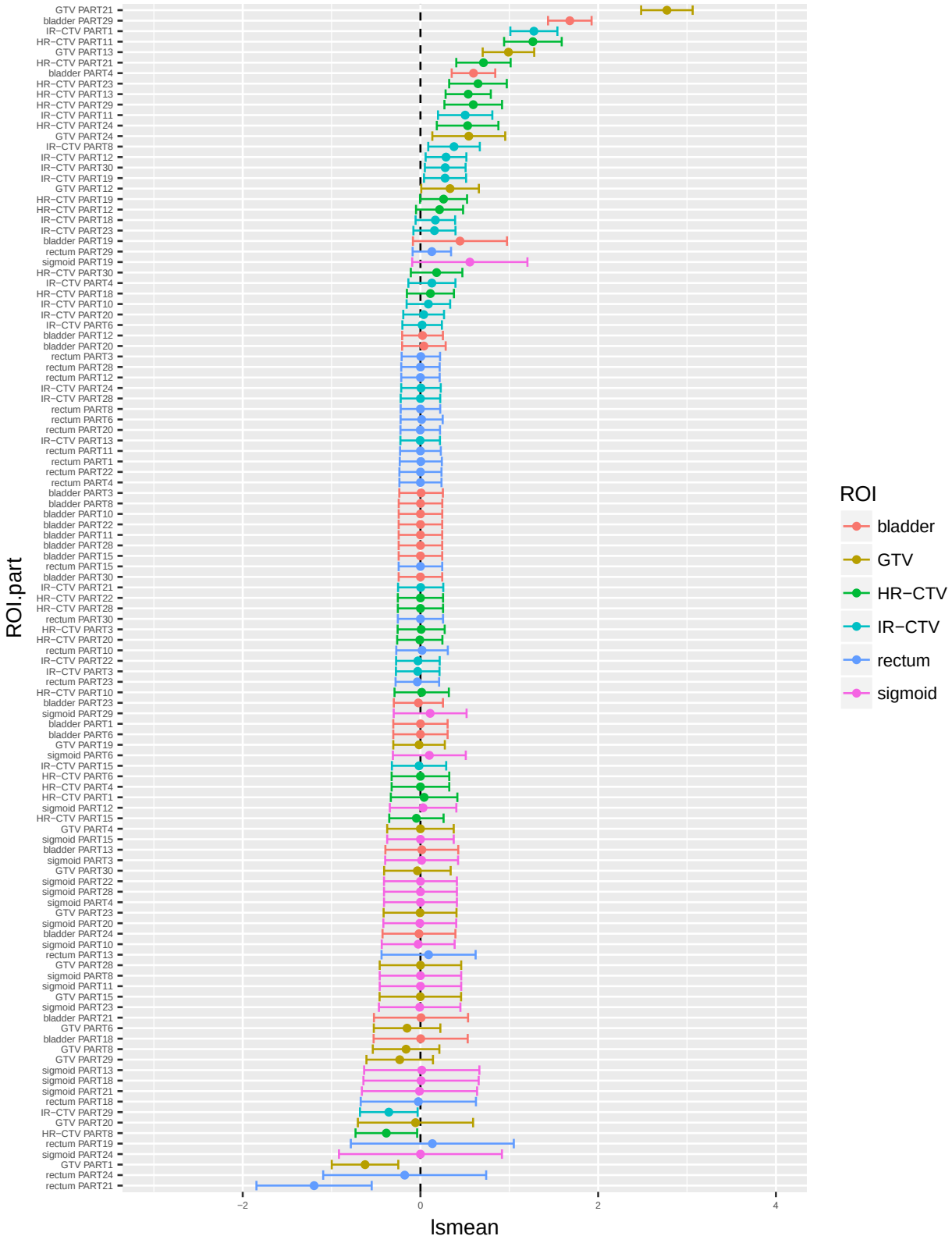


Figure 4.9: Difference of the DICE scores by participant between C2 vs. C1 by ROI. The values are the lsmeans estimates with 95% confidence interval (logit scale) obtained from **modelINTRA.PART** for BT. Participants improve when the lsmean value is positive and the confidence interval does not overlap 0.

Hypothesis tested for C2 vs. C1	*p-value* for EBRT	*p-value* for BT
Does participant significantly improve ($FEtest_{better}$)?	0.00 ***	0.00 ***
Does participant significantly decrease ($FEtest_{worse}$)?	0.75	0.72
Does the improvement depends on ROI.type ($FEtest_{ROI.type}$)?	0.03 *	0.00 ***
Does the improvement depends on inst ($FEtest_{inst}$)?	0.00 ***	0.89
Does the improvement depends on exp ($FEtest_{exp}$)?	0.26	0.48
Does the improvement depends on img ($FEtest_{img}$)?		1.00
Hypothesis tested for C3 vs. C1	*p-value* for EBRT	*p-value* for BT
Does participant significantly improve ($FEtest_{better}$)?	0.00 ***	0.09
Does participant significantly decrease ($FEtest_{worse}$)?	0.16	0.44
Does the improvement depends on ROI.type ($FEtest_{ROI.type}$)?	0.57	0.01 *
Does the improvement depends on inst ($FEtest_{inst}$)?	0.44	0.05
Does the improvement depends on exp ($FEtest_{exp}$)?	0.62	0.02 *
Does the improvement depends on img ($FEtest_{img}$)?		0.40
Hypothesis tested for C3 vs. C2	*p-value* for EBRT	*p-value* for BT
Does participant significantly improve ($FEtest_{better}$)?	0.01 **	1.00
Does participant significantly decrease ($FEtest_{worse}$)?	0.05	0.09
Does the improvement depends on ROI.type ($FEtest_{ROI.type}$)?	0.49	1.00
Does the improvement depends on inst ($FEtest_{inst}$)?	0.39	1.00
Does the improvement depends on exp ($FEtest_{exp}$)?	0.46	1.00
Does the improvement depends on img ($FEtest_{img}$)?		1.00

Table 4.13: Fisher's exact tests to assess the significance of the improvement of the DICE score between the different contouring periods and its association with other effects for EBRT and BT. The answer to the Hypothesis tested is *Yes* when the *p-value* is < 0.05.

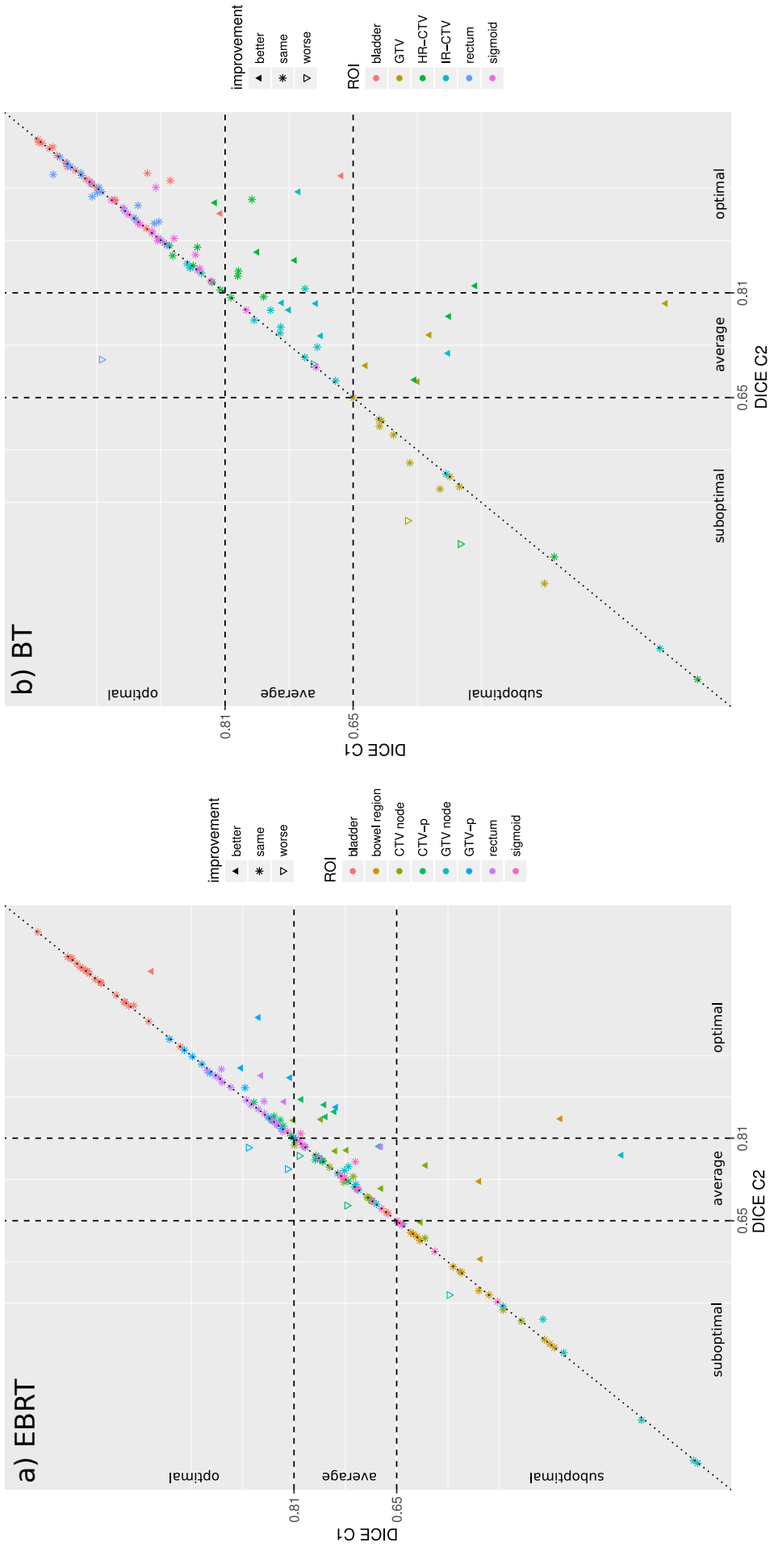


Figure 4.10: The average score for each participant and each ROI for the C2 (x-axis) and C1 (y-axis) for (a) EBRT and (b) BT estimated from the **modelSCORE.PART** and whether the delineation improved or not (the difference is significantly better, equal or worse, from **modelINTRA.PART**, intraobserver variability). Examples (Ex.): Ex. 1.: This participant's DICE index did not vary significantly (same) between C2 vs. C1 for GTV, staying within the suboptimal category. Ex. 2.: This participant's DICE index was in significant detriment (worse) between C2 vs. C1 for rectum, changing from the optimal to the average category. Ex. 3.: This participant's DICE index improved significantly (better) between C2 vs. C1 for IR-CTV, changing from the average to the optimal category (Rivin del Campo et al., 2017).

4.5 Qualitative intraobserver variability

In the EBRT treatment, on one hand the percentage of *correct* contours was significantly better between C2 vs. C1 for posterolateral right in TV. When considering each ROI independently, it was significantly better between C2 vs. C1 for CTV-node and GTV-p, uterus and vagina, at least 20 mm below GTV-p (CTV-p) in the posterior direction, and for CTV-node in the posterolateral left direction. On the other hand, it was significantly worse between C3 vs. C1 and C3 vs. C2 for anterolateral right and inferior in TV, and significantly worse between C3 vs. C2 for posterior and posterolateral left in TV. For independent ROI, it was significantly worse for C3 vs. C2 for anterolateral right for CTV-p and GTV-p, in the posterolateral left direction for CTV-node and for the inferior direction for CTV-p (**Figure 4.11**).

For BT, the percentage of *correct* contours is significantly better between C2 vs. C1 for right and posterolateral right in TV and between C3 vs. C1 for posterolateral right in OAR, while it significantly worse between C3 vs. C1 and C3 vs. C2 for posterolateral right in TV. In the case of the individual ROI, it was significantly worse between C3 vs. C1 in the posterolateral right direction for HR-CTV and between C3 vs. C2 in the same direction for HR-CTV and intermediate risk CTV (IR-CTV) (**Figure 4.12**).

Figure 4.11: Difference of the proportion of correct contours between the different contouring periods. The test of McNemar was used to compare the proportions. Significant difference are indicated by *.

Figure 4.12: Difference of the proportion of correct contours between the different contouring periods. The test of McNemar was used to compare the proportions. Significant differences are indicated by *.

4.6 Satisfaction questionnaire

The scale used for the Organization and Content items went between 1 and 5, with 1 being poor and 5 excellent. The scores for these 20 items for the 20 ODW participants who responded the satisfaction questionnaire of the 32 that submitted contours ranged from 3.95-4.60 with an average of 4.358 (**Table 4.14**).

Participants were also asked whether they would follow another ODW, 80% of the participating clinicians answered positively, and 85% of them would recommend one.

QUESTION TYPE	QUESTIONS	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	N13	N14	N15	N16	N17	N18	N19	N20
BACKGROUND INFORMATION	Gender	2	1	2	2	2	1	1	2	2	2		2	1	2	1	1	1	2	1	2
BACKGROUND INFORMATION	Specialty	RO	RO	RO	RO	RO	RO	RO	RO	RO	RO	RO	RO	RO		RO	RO	RO	RO	RO	RO
BACKGROUND INFORMATION	Num. Years	3	1	10	6	10	3	6	15	3	28	6	19	5		5	13	5	15	6	-
BACKGROUND INFORMATION	Learn about the WS	3	2	3	3	1	3	1	1	3	1	1,3	1	3	3	1	3	1	3	3	4
BACKGROUND INFORMATION	Previous attendance to ESTRO events	0	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
BACKGROUND INFORMATION	Do you perform contouring in cervical cancer?	3,5	5	4	4	4,6	4	4,6	4,6	4,6	4,6	4,6	2,3,5	5	3,5	4,6	4	4,6	4,6	4,6	4,6
BACKGROUND INFORMATION	Do you have experience with contouring in:	3	5	1	1	3	1	1,2,3,4,5,6	3	1	1,2	3	1	5	1,3	5,6	1	5	1	3	1
ORGANIZATION	The goals of this workshop were clearly stated at the beginning	5	5	5	5	4	4	5	5	5	4	4	5	5	4	4	5	4	5	4	5
ORGANIZATION	The supplementary material was	4	4	4	4	4	4	4	4	4	5		4	5	4	4	5	4	5	3	5
ORGANIZATION	The Webex platform for the live presentations was	4	5	5	5	4	2	4	4	4	4	4	4	5	5	4	5	4	5	4	5
ORGANIZATION	The Falcon-Educase contouring tool was	4	4	5	5	4	4	4	3	5	4	4	5	4	3	4	4	1	4	3	5
ORGANIZATION	The workload demands were realistic for this workshop	5	5	5	5	3	3	4	4	4	4	4	4	5	4	4	5	2	5	4	5
ORGANIZATION	The experts' presentation was	4	5	5	5	4	4	5	4	4	5	3	5	5	5	5	5	4	5	4	5
ORGANIZATION	The discussion by the experts and the tutors was	4	4	5	5	4	4	4	1	4	5	3	4	5	4	5	5	4	5	4	5
ORGANIZATION	The language used was clear and understandable and facilitated interaction	3	5	5	5	4	5	4	5	4	5	4	5	5	5	5	5	4	5	4	5
ORGANIZATION	The experts and tutors continuously encouraged communication	4	3	5	5	4	4	4	4	4	4	3	5	5	4	5	5	4	5	5	5
ORGANIZATION	The support from ESTRO and tutors for the delineation exercise was	4	4	5	5	4	4	4	4	5	5	4	5	5	4	5	5	4	5	5	5
ORGANIZATION	The learning experience was equivalent to the experience in face-to-face WS	3	4	4	4	4	4	4	4	4	4	3	4	5	4	5	5	3	4	4	5
ORGANIZATION	The workshop met my expectations	4	4	4	4	4	4	4	4	4	5	3	5	5	5	5	5	3	5	4	5
CONTENTS OF THE WORKSHOP	The WS should allow the participants to obtain a thorough understanding of the contouring guidelines in CC.	4	5	4	4	4	4	4	5	5	5	4	5	5	4	4	5	4	5	5	5
CONTENTS OF THE WORKSHOP	The WS should allow the participants to obtain a good knowledge of contouring in CC and useful imaging modalities to do so.	4	3	4	4	4	4	3	5	5	5	4	5	5	4	4	5	4	5	4	5
CONTENTS OF THE WORKSHOP	The workshop should allow the participants to improve their contouring skills	5	4	5	5	4	5	4	5	5	4	3	5	5	5	4	5	4	5	4	5
CONTENTS OF THE WORKSHOP	2. Provide your overall rating of the quality of the education offered at this WS	4	4	5	5	4	4	4	5	5	5	3	4	5	4	4	5	4	5	4	5
CONTENTS OF THE WORKSHOP	3. Was the information useful and relevant to your work and practice techniques?	4	4	5	5	4	4	4	3	5	4	3	5	5	4	4	5	4	5	3	4
CONTENTS OF THE WORKSHOP	4. Do you feel that the presented information was well balanced and supported by adequate evidence?	5	5	4	4	4	4	4	5	5	5	4	5	5	5	4	5	4	5	5	4
CONTENTS OF THE WORKSHOP	5. Did the programme allow adequate time for discussion and questions?	4	5	5	5	4	5	3	4	4	4		4	5	5	4	5	2	5	4	3
CONTENTS OF THE WORKSHOP	6. How would you rate the management and organization of this WS?	3	4	5	5	3	4	4	4	5	5	4	4	5	4	4	5	4	5	4	5

Table 4.14: Post-ODW satisfaction questionnaire. N: participant number. Gender: 1: male; 2: female. RO: Radiation Oncologist. Learn about the workshop: 1: RAIDs project manager and/ or WP-4.2 leader IGR; 2: Department director; 3: Colleagues. Previous attendance to ESTRO events: 0: no; 1: yes. Do you perform contouring in cervical cancer?: 1: Never; 2: Sometimes –specify: 3: Yes, for EBRT under supervision; 4: Yes, for EBRT without supervision; 5: Yes, for BT under supervision; 6: Yes, for BT without supervision. Do you have experience with contouring in: 1: CT-images only; 2: MRI images only; 3: Matched CT and MRI images; 4: PET-images; 5: Multimodality (PET-CT-MRI), Ultrasound images. The remaining items were scored on a scale from 1-5, 1 being poor and 5 being excellent, or for some items under *contents*: 1 being no and 5 being yes, with the numbers in between being: relatively.

5 Discussion

Contents

First and foremost, it is essential to underscore the innovative use of this online delineation workshop (ODW) to evaluate delineation skills as a *dummy run* within a multicentre trial. This is the first time this kind of ODW has been used for this purpose within this context (Grau Eriksen et al., 2014). Recently, other clinical trials have also employed an ODW following this format as part of their quality assurance programmes. An example is HYPO-G-01, which has promising initial results submitted as an abstract to the next European Society for Radiotherapy & Oncology (ESTRO) meeting (ESTRO 37, (*HYPOG-01, https://www.gustaveroussy.fr/en/node/3646* 2017)). Other authors have recommended including training programmes in Radiation Therapy Quality Assurance (RTQA) protocols (Fokas et al., 2015). The results presented support the feasibility of the ODW as a contouring *dummy run* within a multicentre trial and its capacity to identify centres with baseline and subsequent average to optimal contours which are prepared to include patients, while providing an educational tool for others. The additional contribution of this study is that it also presents the participants' point of view, which in regard of the results of the post-ODW

satisfaction questionnaire is quite favourable.

A comprehensive follow-up model was developed within a publication led by the European Organisation for Research and Treatment of Cancer (EORTC) quality assurance strategic committee and team, to use individual patient data to establish a possible impact of RTQA levels on patient outcome (Weber et al., 2014). They found a clear correlation between survival (both quality-of-life-adjusted and recurrence-free) which augmented along with the RTQA levels. Notably, with a level 4 RTQA programme there was a gain of 1.8 months of QALY, compared with a level 2 RTQA. This is much higher than the QALY which is usually gained in prospective randomised radiotherapy trials. This leads to a very interesting statement: considering the costs of a tumour recurrence, higher RTQA levels, though costly, actually induced significant savings (Weber et al., 2014).

Specifically, within the context of cervical cancer, the RTQA programme performed within the Induction Chemotherapy Plus Chemoradiation as First Line Treatment for Locally Advanced Cervical Cancer (INTERLACE) trial resulted in the development of a contouring atlas (Eminowicz et al., 2016a). As previously mentioned, when this atlas was used within INTERLACE, the average proportion of protocol compliant delineations for primary clinical target volume (CTV), nodal elective volume (CTV-node), bladder and rectum increased from 1.8 to 2.7 (difference of 0.9; 95% confidence interval, 0.3-1.5; $p\text{-}value = 0.003$). Its use also lowered interobserver contouring variation within INTERLACE (Eminowicz et al., 2016a).

An issue of utmost importance in delineation studies is the choice of the metrics used for the comparison. In our experience, it was the the DICE score, however, many different metrics exist, as reported by Taha and Hanbury, (2015). DICE is the metric used most often to evaluate three-dimensional (3D) medical imaging segmentations, especially to determine reproducibility. It is an overlap metric, as well as the Jaccard index, the true and false positive rates (not commonly used, due to their sensibility to the size of the segment) and the global consistency error. Of note, the correspondence between DICE and distance based metrics is reduced when the overlap is smaller, which is logical as DICE only takes into consideration the voxels which are in the region that overlaps, as opposed to distance based metrics (Taha and Hanbury, 2015). For this reason, DICE is less dependable for small volumes, explaining low median DICE scores in studies evaluating small volumes, such as Chung et al., (2015). Thus, adding a distance based metric to our analysis may have resulted in more robust results. With this intention, the contour error distance tool was developed within Fellowship in Anatomical deLineation and CONtouring (FALCON) EduCase^{TM} by

petition in the context of this thesis. However, it was more within views of an objective qualitative interpretation of the contouring variability, as it did not supply exact quantitative results.

5.1 Interobserver validation of the ODW in LACC

It has been both numerically and graphically described in the literature, even specifically for cervical cancer, that the largest contouring uncertainties are on the cranial and caudal slices of a volume (Petrič et al., 2013; Hyeon Joo et al., 2017). This fully agrees with our results for the quantitative interobserver analysis (**Figure 4.2**). The most particular pattern belongs to the bowel, since the expert contour followed individual bowel loops and instead most participants contoured a bowel bag, although both contouring techniques are valid (Banerjee et al., 2013). Barillot et al., (2014) highlight how clear contouring definitions should be established for bowel in patients receiving intensity modulated radiotherapy (IMRT) to the pelvis, especially in cases of pelvic nodal irradiation.

The main reason why the evaluation of interobserver variability was the first primary endpoint was it allowed assessment of the overall improvement/detriment of all participants in the ODW between them, instead of considering only individual variability versus the expert contour. This is of uttermost importance since the expert contour affects the comparison of region of interest (ROI), and most often has certain flaws (Vinod et al., 2016; Vinod et al., 2017). Not surprisingly, the interobserver comparisons for both external beam radiotherapy (EBRT) and brachytherapy (BT) showed overall more improvement between guideline contouring (C2) and baseline contouring (C1) than between final contouring (C3) and C1, and mostly the worse results were between C3 and C2. This seems to indicate that participants acquired contouring skills after the guideline presentation, and just partially retained this knowledge 1.5-2 months later. However, a positive point is that the intraobserver analysis showed that only improvements were significant between different contouring periods.

As for the results for interobserver variability considering experience in EBRT, experienced specialists had significantly worse results than less experienced specialists and senior residents for radiologically pathological lymph nodes to boost (GTV-node). But this finding should be interpreted with caution. A borderline significant paraaortic lymph node (though *no positive paraaortic lymph nodes* was specified in the clinical case) and a suspicious left inguinal node were visible on the EBRT computed tomography (CT) image set. The latter was considered inflammatory by CHM in the live sessions. Thus, interestingly, this may highlight

that less experienced specialists and senior residents fully integrated all the clinical information provided when contouring. As could be expected, less experienced specialists, with more experience, did better than junior residents for GTV-node and than junior and senior residents for sigmoid, as senior residents did better than junior residents for GTV-node (all results were significant). Unexpectedly, junior residents had significantly better results than senior residents for cervix, parametria and vaginal gross disease (GTV-p). There was only one significant difference related to experience for BT: experienced specialists contoured better the sigmoid than senior residents. Interestingly, results from the An intErnational study on MRI guided BRachytherapy in locally Advaced CErvical cancer (EMBRACE) *dummy run* may somewhat explain this, as they found that the organ at risk with the most divergence was the sigmoid, in 7 of the 28 centres (Kirisits et al., 2015).

A qualitative analysis of the interobserver variability was also contemplated within the first primary objective. When addressing this analysis from a statistical point of view, it was not feasible due to several reasons. First, the data from the qualitative analysis are by nature discrete. The statistical model **modelINTER.PART** we proposed for the interobserver variability in the quantitative analysis cannot be applied on discrete data. Second, the qualitative classification of contours with the three modalities **correct**, **incorrect**, **very incorrect** has been performed for eight different directions of space. Therefore, as opposed to the DICE score, there is not a unique qualitative score that could be used to evaluate the interobserver variability. Third, even if this score would exist, it should be inferred automatically from the contours that have been submitted, while the qualitative classification has been performed from a visual inspection, done manually. Considering this last task would be very long and tedious, it would be very difficult to gather the necessary data. Of note, Petric et al., (2008) did manage to perform a qualitative analysis of interobserver variability for a BT treatment. They measured manually the distances between the centre of the uterine tandem to the high risk CTV (HR-CTV) contour of both observers in the study, in the eight directions of space. The difference was that they only did this for one ROI, as opposed to the 14 ROI analysed in this study, and for only two observers instead of up to 30 for C1 for the BT treatment in our experience.

Another possible way to analyse qualitative interobserver variability may have been by inclusion of anatomic regions, as performed by Fairchild et al., (2014) and Hyeon Joo et al., (2017). Unfortunately, in the context of the ODW, the workshop structure would need to be modified for this. When it was performed, all necessary regions included were specified, except for the lymph node stations. Similarly to the (Hyeon Joo et al., 2017), a frequent

difficulty shown by the participants when contouring the CTV-node was including the 7 mm margin around the pelvic vessels, although specific recommendations were provided (Hyeon Joo et al., 2017; Taylor et al., 2007).

5.2 Intraobserver validation of the ODW in LACC

The most interesting aspect of the quantitative results for intraobserver variability is that only the improvements of the participant contours for each clinician were significant between contouring periods (**Table 4.13**). For C2 vs. C1 the improvement depended on ROI type for the contours for both treatments and also on institution for the EBRT treatment contours. Conversely from the EBRT treatment, for BT overall treatment contours there was no significant improvement between C3 vs. C1 and C3 vs. C2. On the contrary of what could be assumed, the use of magnetic resonance imaging (MRI) imaging for planning of BT treatments had no significant impact from the intraobserver point of view.

In the literature, a qualitative interobserver analysis did not show differences in the 8 directions of space for HR-CTV contours (Petric et al., 2008). The experience performed for this thesis evidenced qualitative intraobserver differences for correct contours which were significant for certain directions (better or worse) for EBRT. However, there was no obvious explanation to this phenomenon, and these differences did not seem to have a clear potential clinical impact. Otherwise, for BT, the significant improvement towards the right and posterolateral right for target volume (TV) in C2 vs. C1 most probably was because the left parametrial invasion made participants focus more on the left portion of the TV than on the right during C1, and they improved after the guideline session. But unfortunately this improvement was shortlived, and in C3 they went back to their old ways, doing significantly worse for posterolateral right TV in C3 vs. C1 and C2.

Centres were identified with participants with suboptimal baseline contours which did not significantly improve and change categories (to average or optimal) in subsequent contouring periods. Clinicians from these centres were invited to follow-up specific, onsite training, in standard radiochemotherapy for locally advanced cervical cancer (LACC), at our centre.

5.3 Evaluation of the teaching methodology of the ODW in LACC

The results of the submitted student satisfaction questionnaires were excellent. The best satisfaction results were reflected for reaching the goals stated for the ODW, the expert presentation, the support offered by ESTRO and the tutors, the language used, and in agreement that the information presented was well balanced and supported by the evidence. When asked whether they considered the learning experience as equivalent to face-to-face the average response was 4.05 on a scale of 1-5, which shows a good acceptance of the virtual nature of the ODW. The average response for the overall rating of the ODW was even better, 4.4. Thus, the clinicians clearly supported not only this innovative workshop format, but they also valued well its organisation, support and content.

The average answer of participants was 3.95 on a scale of 1-5 for evaluation of the FALCON EduCase^{TM} contouring platform, reflecting a good understanding of its use. A previous publication has recommended each centre use their own treatment planning system (TPS) for contouring of *dummy runs* to ease the use of the knowledge acquired in routine clinical practice (Clark et al., 2009). Though this approach may be useful, especially for the implementation of new modern radiotherapy techniques in an institution within a multicentre trial, from a practical standpoint it requires more resources to transfer image sets and contours. Conversely, FALCON EduCase^{TM} is a fully virtual online contouring platform, which does not require downloading of any modules to be functional, and is accessible from anywhere, anytime. Its accessibility may also favour compliance with the *dummy run* exercises.

Another strong point of the ODW blended learning experience is it allowed immediate interaction between participants and tutors, and timely interaction with the teaching faculty. As the ODW was performed before opening Rational molecular Assessments and Innovative Drug Selection (RAIDs) centres, certain ambiguities in the RAIDs contouring protocol were detected by participant-tutor interaction. This resulted in the corresponding modifications of the RAIDs contouring protocol. Similarly, Lo et al., (2014) have found that a large part of the recommendations to adapt delineations in their contouring study for stereotactic body radiation therapy (SBRT) in lung cancer were due to either to nonadherence to the contouring guidelines, or to ambiguities in their interpretation. As in our experience, a *dummy run* in the EORTC 22043-30041 trial in postoperative prostate radiotherapy +/- androgen deprivation therapy achieved improvements in the trial contouring protocol by pointing out its weak points before patient accrual (Fenton et al., 2013). This indicates the

need for a clear hands on explanation of delineation guidelines and a continuous review of the contours to motivate discussions which may bring forth frequent areas of divergence (Lo et al., 2014).

From the point of view of the e-learning educational experience, an important aspect of the ODW is that it allows a self-directed path. Each participant may choose to attend live online sessions, within a blended learning model (with support and interaction with tutors) and/or follow the session recordings. This flexibility is essential in this context, as the ODW may be adapted to the physicians' heavy workload (Booth et al., 2009). But unfortunately this was not effective in all cases, 14 of the initial 46 enrolled participants did not submit contours. Although, it is to be noted, that the 20 participants who did submit contours and answered the satisfaction questionnaire considered that the workload demands of the ODW were realistic (average score of 4.2 on a scale of 1-5).

As previously mentioned, this ODW followed the structure of ESTRO FALCON ODW. ESTRO has been providing onsite delineation workshops since 2009, and ODW since 2012. Not only has ESTRO taken action to provide the necessary onsite and online training courses in contouring for the radiotherapy community, but so has the Royal College of Radiologists (RCR) (Eminowicz et al., 2016b).

5.4 Analysis of clinician contouring on MRI for BT planning

Evidence has shown that in MRI image guided brachytherapy (IGBT), MRI especially allows better visualisation of the vagina and uterus than of the rectum or bladder (Dimopoulos et al., 2006). This may very well be an explanation for the interobserver quantitative improvement in HR-CTV and intermediate risk CTV (IR-CTV) for C2 and C3 vs. C1, as opposed to a detriment for the bladder (C3 vs. C2) (**Table 4.5**).

It is also interesting to note the significant interobserver quantitative improvement for contouring of HR-CTV for centres doing MRI-IGBT vs. those that do not, showing the impact of specific institutional training in MRI-based contouring (**Figure 4.7**). An inter-observer contouring comparison has been performed between two centres with a tradition in MRI-based contouring. In that publication, only mean volumes of IR-CTV differed significantly between institutions, but with no significant differences when considering conformity indices , as could be expected from the world renowned participating institutions (Dimopou-

los et al., 2009). Thus, they showed that the use of Groupe Européen de Curiethérapie – European Society for Radiotherapy and Oncology (Gyn GEC-ESTRO) guidelines for IGBT contouring yields acceptable interobserver variability. In effect, training in contouring and the use of specific contouring guidelines may reduce systematic contouring errors (Dimopoulos et al., 2009).

5.5 Preworkshop questionnaire

Much practical information of the radiotherapy techniques practised within the participating RAIDs centres was gathered from the pre-workshop questionnaire. In light of the publication by Mazeron et al., (2017), special attention may be given to the item: BT prescription.

Mazeron et al., (2017) have evaluated the relevance of reporting the dose to point A as in classical BT treatments in the IGBT era. They only found a significant relationship between the dose to point A with total reference air kerma (TRAK) and one IGBT treatment volume: as a surrogate of the D90 (the dose to 90% of the volume) HR-CTV. The TRAK is a function of the volume irradiated. The dose to point A increases with the TRAK, as well as with the decrease of D90 HR-CTV, by a formula involving the HR-CTV volume for the latter. They did not find a direct significant correlation between point A doses and D90 HR-CTV, nor with local control, leading them to clearly question the utility of reporting point A doses for IGBT treatments. But the International Commission for Radiation Units and measurements. (ICRU) still advocates routine point A reporting in IGBT (Mazeron et al., 2017) .

In our study, 15 out of 17 participating RAIDs centres performed IGBT. Of them, only three centres still prescribed to point A, as well as to the HR-CTV. Only one of these three centres did MRI-based IGBT. However, our questionnaire did not reflect whether the remaining 12 centres doing IGBT included the dose to point A in their treatment reports.

Of note, in the EMBRACE II study, a supplementary planning aim to the IR-CTV and HR-CTV aims was point A $>$ 65 Gy. Its objective was to guarantee a minimum dose in small tumours, where contouring uncertainty may cause insufficient tumour treatment coverage. The Mazeron et al., (2017) patient series showed no difference in 3-year local control for small lesions (HR-CTV $<$ 3 cm) between those that received a dose to point A $\geq$65 Gy or $<$ 65 Gy. Adding this planning aim to these patients would have incurred in a higher dose to the organs at risk (OAR) with no clear clinical benefit, for the time being (Mazeron et al.,

2017).

5.6 Limitations of this study

The first limitations encountered during this study were organisational. It proved difficult to locate radiotherapy professionals since the ODW was held before the RAIDs trial opened in clinical centres. Furthermore, the first ODW took place in June-July. In this context, many clinicians could not participate or only could attend some sessions. Fewer contour sets were submitted during July (C2), and even less in August-September (C3). Also, aversion of participating centres to complete radiotherapy credentialing programmes has been previously reported, and mostly attributed to the time needed to perform these activities (Weber et al., 2012). This led to only 13 contours being submitted for C3, limiting statistical significance and with a less representative population.

Within the submitted contour sets, not all OAR were contoured for all three contouring attempts. This is in line with the individual case review (ICR) performed within a previous study, the EORTC 22033–26033/CE5 phase III randomised low grade glioma trial (Fairchild et al., 2012c). In an average of 1/4 of the cases (range 5-72%) required OAR contours were absent, and almost 1/3 of the present OAR contours were not correctly contoured (Fairchild et al., 2012c).

Another limitation, as somewhat previously mentioned, was the use of the DICE index. It was the only contouring conformity index available as FALCON EduCase^{TM} output at the time. This index is less reliable in small ROI volumes. Examples in our study would be GTV-node, gross tumor volume (GTV) or sigmoid, showing lower concordance because slight differences in delineation have more impact on the score. Conversely, in very large ROI volumes, as CTV-node or bowel, it seems to lack the sensitivity to recognise divergences from the reference contour (**Figure 4.2**) (Breunig et al., 2012; Esthappan et al., 2011). This phenomenon is due to the duplication of the overlapping volume, which may falsely show considerable agreement in these large ROI. But this index is very simple to calculate, making it is the most used index in automatic segmentation studies (Hoang Duc et al., 2015). The DICE index may also be converted into other overlap indexes by using certain ratios (Fotina et al., 2012).

An ultimate limitation to performing a more exhaustive RTQA programme was trial monitoring of the electronic case report forms (eCRFs) and the financing allotted. Initial plans were to perform a more complete RTQA assessment, including an ICR of the first

patient included and a patient chosen at random from each centre. This review would have included the clinical case, delineation and treatment planning. Several difficulties were encountered. Firstly, centres had many difficulties filling in correctly the radiotherapy sections of the eCRFs. They were slightly modified by the RAIDs RTQA group to improve their comprehensibility, but due to coordination issues many of these modifications and specifications were not incorporated into the eCRFs. Fortunately, the site visits performed by the study monitors identified different interpretations between centres of the radiotherapy eCRFs, which helped improve the quality of this data, as evoked by Haworth et al., (2009). This led to an extensive collaboration between the RAIDs RTQA group and the study monitors to achieve accurate, interpretable radiotherapy data by the end of the 5 year RAIDs project. Thus a real-time or even coetaneous ICR could not be performed. Secondly, as previously mentioned, obtaining collaboration from radiotherapy professionals from participating RAIDs centres proved complicated for the ODW and surely would have been even more difficult for an ICR. Lastly, the budget for RTQA covered the realisation and posterior analysis of the contouring *dummy run* within the ODW, with no excess funding for the creation of a server to import the diagnostic image sets, radiotherapy treatment image sets, contours and plans.

These limitations are not unique to this study. Poortmans et al., (2006) have reported the results from the *dummy run* and the ICR for the EORTC 22922/10925 addressing the role of radiotherapy of the internal mammary and supraclavicular lymphatic chains. Forty-one of the 45 participating institutes participated in the *dummy run* and only 20 (less than 50% of those which completed the *dummy run*) did the ICR. The positive side is that the institutes which completed the *dummy run* included 93% of the study patients, and those which did the ICR included 76%. Thus, the most motivated centres in the study correctly followed the RTQA programme (Poortmans et al., 2006). In the current study, the centres which participated in the *dummy run* accrued 66% of all RAIDs patients. The RTQA programme of the Japan Clinical Oncology Group (JCOG) 0202 study (cisplatin-etoposide versus cisplatin-irinotecan in consolidation chemotherapy for early stage small cell lung cancer) consisted in an ICR of all accrued patients, without a previous *dummy run* (Sanuki-Fujimoto et al., 2009). They found a gradual reduction of unacceptable variations, which were mostly within the first 3 patients. Their explanation was the feedback to centres on protocol compliance, especially considering that no *dummy run* was performed before patient accrual (Sanuki-Fujimoto et al., 2009). This clearly highlights the paramount importance of performing a *dummy run*, even when an exhaustive ICR is performed. Another Japanese publication

presented the RTQA results of the ICR of the (JAROG0401/JROSG04-2) study in early stage cervical cancer, aiming to evaluate the efficacy and toxicity of radical EBRT followed by high dose rate (HDR) BT (Toita et al., 2009). The BT was administered by a two-dimensional (2D) technique, and 17% of the ICR showed a deviation of the protocol with regard to the determination of the point A. The authors recognise this 2D technique is outdated and manifest the need for a more stringent RTQA programme in future cervical cancer trials, including a *dummy run*.

Budget issues for RTQA within multicentre trials have been well described in previous publications. Thus several, in line with our study, have only performed a *dummy run*. An example is the Barillot et al., (2014) study on postoperative IMRT for endometrial cancer. Poortmans et al., (2005) state that the increasing complexity of RTQA programmes due to modern radiotherapy techniques such as IMRT and SBRT convey a larger workload and an increased cost. In the present study, both IMRT and image guided radiotherapy (IGRT) could be used, as well as classic radiation techniques like 3D radiotherapy and two dimensional BT.

5.7 Future perspectives

Several efforts are being made to improve and homogenise RTQA programmes, so as to reach the perfect balance between thoroughness and efficiency. As recently as 2014, the National Clinical Trials Network (NCTN) has been created by the National Cancer Institute (NCI) to coordinate all intergroup Phase III clinical trials. Within the NCTN a specific group is in charge of imaging and RTQA: the Imaging and Radiation Oncology Core Group (IROC) (Fitzgerald, 2013). The key role of IROC is to implement an efficient and effective RTQA workflow, by integrating standard operating procedures for imaging and radiotherapy dataset transfer, as well as providing initial site evaluations and credentialising, protocol support, transfer of electronic data, data management and ICR (Fitzgerald, 2013). The goal of IROC is to streamline the process, allowing for a uniform and rigorous RTQA programme across the NCTN (Fitzgerald, 2013).

Within the EORTC much of the aforementioned processes may be performed using the Visualization and Organization of Data for Cancer Analysis (VODCA) system (Weber et al., 2011). This system even creates dose-response models, allowing evaluation of the relationship between results and RTQA compliance in trials. This is essential to present the validity of these results in light of high quality radiotherapy (Weber et al., 2011).

Future directions in RTQA must incorporate common RTQA baseline centre credentialising, allowing centres participating in several trials to submit only specific credentialising data for each trial, thus decreasing their workload (Miles and Venables, 2012). This will also permit efforts to be centred on audits and updating of the status of the trial centres (Miles and Venables, 2012).

Two multicentre trials considered the main RTQA method should be the *dummy run*, and not a prospective interventional ICR of all patients (Fairchild et al., 2012a). In the TROG 0202 head and neck study they felt that the latter could not rectify planning deviations within an acceptable timeframe (L. J. Peters et al., 2010). In the case of the EORTC 20884 lymphoma study they believed the workload was too cumbersome (Aleman et al., 2005). Clearly, automatising of ICR within a digital platform such as VODCA allowing for real-time case reviewing may change these conclusions. Innovative approaches to automating RTQA include a pilot study applying a knowledge base built from nine delineations from 29 head and neck cancer treatment plans (Altman et al., 2015). The base calculated several metrics from the plans, such as the shape, size, position in relation to other structures, etc. and determined heuristically derived rules. They analysed nine more plans with 42 contouring errors, and the knowledge base identified 40 of these errors, along with 9 false positive results (Altman et al., 2015). Potentially, this could be applied within digital RTQA platforms.

A final perspective to be performed within the scope of this thesis will be an analysis of the eCRFs to correlate the RT treatments administered with morbidity and response. It will also allow for a general description of RT practices within different European centres. As previously mentioned, curated eCRFs data has only become available in late 2017, thus this will be performed during the following months. Depending on the quality of the eCRFs data, either an initial analysis will be performed presenting local control and acute morbidity, followed by a second analysis evaluating long term local control, survival and chronic morbidity, or if long term follow up data is unavailable, only the first analysis will be possible. The data in the eCRFs include the overall treatment time, and if it exceeds 55 days its possible impact on local control shall be evaluated (Haie-Meder et al., 2010b; Tanderup et al., 2016). Unfortunately these issues are not infrequent, an example is Fairchild et al., (2012c). In that study on low grade glioma, case report forms often were incomplete or contradicted the results of the digital review. Since unfortunately in this study a digital review of the cases was not performed, the monitoring of the eCRFs data is relied upon.

6
Conclusiones

1. En general, se observó sobre todo una mejoría significativa en la variabilidad interobservadora e intraobservadora cuantitativa entre contornos guía (C2) y basales (C1) para los dos tratamientos, y la única diferencia cualitativa notable fue una mejoría de la delineación de volúmenes blanco entre estos periodos para braquiterapia (BQT).

2. Los resultados del cuestionario de satisfacción han dejado patente que los participantes consideran la metodología de enseñanza del taller de delineación en línea (TDE) muy apta.

3. El análisis cuantitativo de la variabilidad interobservadora entre los centros que usan de manera rutinaria la resonancia magnética para la delineación de tratamientos de BQT y los que no, mostró mejores resultados en contorneo del high risk CTV (HR-CTV) en los centros acostumbrados a esta técnica.

Concluimos que el TDE ha sido validado para el asesoramiento inicial de la delineación en centros geográficamente distantes, permitiendo un control de calidad inicial de un ensayo multicéntrico en cáncer de cérvix localmente avanzado (CCLA), asimismo ofreciendo una formación inicial en delineación. Sin embargo, en el futuro los esfuerzos deben dirigirse a mejorar esta formación, especialmente en cuanto al efecto de la enseñanza a largo plazo (contornos finales (C3)). Los centros con participantes que necesitan mejorar deben tener la posibilidad de continuar formándose, siguiendo una secuencia óptima de métodos en línea, en persona o una combinación de ambos.

7
Conclusions

1. Overall, there was mostly a significant improvement for quantitave interobserver and intraobserver variability between guideline contouring (C2) and baseline contouring (C1) for both treatments, and the only notable qualitative difference was an improvement in target volume (TV) delineation between these periods for brachytherapy (BT).

2. The results of the satisfaction questionnaire have clearly proven that participants highly appreciate the teaching methodology of the online delineation workshop (ODW).

3. The quantitative analysis of the interobserver variation between centres using magnetic resonance imaging (MRI) routinely for BT planning and those that do not showed better performance in contouring for the HR-CTV for centres accustomed to this technique.

Thus, the ODW has been validated for initial assessment of delineation in geographically distant centres, allowing baseline quality control for a multicentre trial in locally advanced cervical cancer (LACC), as well as offering initial training in delineation. However, future directions should insist on improvement of this training, especially with respect to the long term teaching effect (final contouring (C3)). Centres with participants requiring improvement should be offered further training, following the optimal sequence of online, onsite or blended approaches.

8 Article

Contents

8.1 Assessment of the novel online delineation workshop dummy run approach using FALCON within a European multicentre trial in cervical cancer (RAIDs)

This is the article published within this thesis (Rivin del Campo et al., 2017).

Radiotherapy and Oncology 124 (2017) 130–138

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

FALCON

Assessment of the novel online delineation workshop dummy run approach using FALCON within a European multicentre trial in cervical cancer (RAIDs)

Eleonor Rivin del Campo [a,*], Sofia Rivera [a], María Martínez-Paredes [b], Philippe Hupé [c,d,e,f,g], Andrea Slocker Escarpa [h], Isabelle Borget [i,j], Renaud Mazeron [a], Suzy Scholl [k], Amalia Palacios Eito [l], Christine Haie-Meder [a], Cyrus Chargari [a], Eric Deutsch [a]

[a] Department of Radiation Oncology, Gustave Roussy Cancer Campus, Villejuif, France; [b] Department of Radiology and Medical Physics, Medical School, University of Cordoba, Spain; [c] Institut Curie; [d] PSL Research University; [e] Inserm, U900, Paris; [f] Mines Paris Tech, Fontainebleau; [g] CNRS, UMR144, Paris; [h] Department of Radiation Oncology, IUCT Oncopole, Toulouse; [i] Service de Biostatistique et d'Epidemiologie, Gustave Roussy; [j] INSERM U1018, CESP, Université Paris-Sud, Université Paris-Saclay, Villejuif; [k] Department of Medical Oncology, Institut Curie, Paris, France; [l] Department of Radiation Oncology, Reina Sofia University Hospital, Cordoba, Spain

ARTICLE INFO

Article history:
Received 29 January 2017
Received in revised form 10 April 2017
Accepted 4 May 2017
Available online 19 May 2017

Keywords:
Interobserver variability
Intraobserver variability
Quality assurance
Cervical cancer
Brachytherapy
e-Learning

ABSTRACT

Background and purpose: Online delineation workshops (ODW) permit training of geographically dispersed participants. The purpose is to evaluate the methodology of an ODW using FALCON to harmonize delineation within a European multicentre trial on locally advanced cervical cancer (LACC).

Material and methods: Two ODW included 46 clinicians (14 centres). Clinicians completed baseline (C1), guideline (C2) and final contours (C3) for external beam radiotherapy (EBRT) and brachytherapy (BT) for LACC. Interobserver and intraobserver variability was evaluated quantitatively (using the DICE index) and qualitatively compared to expert contours.

Results: Nine clinicians submitted for EBRT and BT for C1–C3. Thirty-two sent any contour. Interobserver quantitative comparisons for EBRT showed significant improvement for C2 vs. C1 for bowel, CTV node, CTV-p and GTV node with significant detriment for GTV node (C3 vs. C1; C2), CTV-p (C3 vs. C2) and bowel (C3 vs. C2), showing in general an improvement in C2 vs. C1, with a detriment in C3 vs. C2 for two target volumes and an organ at risk. For BT there was significant improvement for C2 vs. C1 for bladder, GTV, HR-CTV and IR-CTV, with significant detriment for bladder (C3 vs. C2), thus overall improvement in C2 vs. C1, with only a detriment in C3 vs. C2 for bladder. Centres using MRI imaging for BT contouring did significantly better in the BT case for HR-CTV than those which used other techniques (C2 vs. C1: $p < 0.005$; C3 vs. C1: $p = 0.02$). Intraobserver quantitative comparisons showed significant improvement contouring a region of interest between C2 vs. C1, C3 vs. C1 and C3 vs. C2 for EBRT and between C2 and C1 for BT.

Conclusions: ODW offer training, initial contouring harmonization and allow assessment of centres.

Much has evolved since the first contouring dummy run including distant centres within a multicentre trial, which used CT hard copies [1]. As described in 1995, online education allows participative medical training for geographically dispersed students [2]. Flexibility, essential within e-learning, especially for medical professionals, defined as 'learner control', offers self-task management [3]. Student outcome evaluation is also important, though few report objective internal testing to validate web-based learning tools as a primary outcome [4–7].

Radiotherapy quality assurance has become key to ensure interpretable results within multicentre trials, especially after reports have shown the influence of contouring on patient outcomes [8–11]. Hence the phase III trial of concurrent cisplatin and tirapazamine in head and neck cancer in which when radiotherapy compliance was analysed, a significant reduction of 2 year overall survival and locoregional control was observed when treatment plans were largely deviated from protocol [8].

Proper delineation of target volumes (TV) and organs at risk (OAR) is crucial, allowing optimal oncological treatment and better knowledge of the dose received by surrounding healthy tissue. Thus, several studies have evaluated interobserver and sometimes intraobserver variability between contours [12–15]. Two recent

* Corresponding author at: Department of Radiation Oncology, Gustave Roussy Cancer Campus, 114 Rue Edouard, Vaillant, 94805 Villejuif, France.
E-mail address: Eleonor.RIVINDELCAMPO@gustaveroussy.fr (E. Rivin del Campo).

http://dx.doi.org/10.1016/j.radonc.2017.05.008
0167-8140/

reviews addressed this issue, one proposing reporting items for these studies, which this paper will adhere to [16,17]. In locally advanced cervical cancer (LACC) this variability acquires even higher significance. Recent advances in External Beam Radiotherapy (EBRT) and Brachytherapy (BT), namely image guided brachytherapy (IGBT), have shown 3 year local control rates of 92% (tumours > 5 cm) and 98% (tumours 2–5 cm) [18]. This was achieved by applying the Gynaecological GEC-ESTRO (Groupe Européen de Curiethérapie – European Society for Radiotherapy & Oncology) recommendations to the high risk clinical TV (HR-CTV) and dose volume constraints for OAR [19].

The purpose of this study is to validate the methodology of an online delineation workshop (ODW) within a European multicentre prospective study in LACC (Rational molecular Assessments and Innovative Drug Selection: RAIDs), which includes 22 European clinical centres including Eastern and Western Europe [20]. To this aim, participant contours in different periods were reviewed, as well as the participants' personal perception of the knowledge acquired.

Materials and methods

Before the ODW a general questionnaire about LACC radiotherapy was sent to RAIDs centres for input on their practice (Table 1).

ODW structure

Two to four participants from each centre (proportional to the gynaecological team) were enroled in an ODW in LACC, exceeding its capacity, thus two ODW were planned. A technical partnership was established with ESTRO. The methodology was similar to that used in FALCON (Fellowship in Anatomical deLineation and CONtouring) ESTRO ODW [21]. Live presentations were via WebEx and contouring was done using the FALCON EduCase™ contouring platform.

Training was given by an expert, CHM, with one tutor per 10 clinicians. Tutors were radiation oncologists with experience in LACC, trained to use FALCON EduCase™. Live sessions were completed in 3 weeks and participants delineated EBRT (on Computed Tomography: CT) and subsequent BT (on Magnetic Resonance Imaging: MRI) image sets for the same clinical case. The case and image sets with expert contours were chosen with CHM, from the ESTRO FALCON EduCase™ contouring library.

The ODW were held on June–July 2013 and January 2014, respectively, with an identical structure. The first two live sessions were presented by tutors.

- Session 1 exposed FALCON EduCase™ and the clinical case. Participants were informed (orally and in writing) that their contours would be in a study evaluating the ODW, requesting their conformity, which was not revoked. Clinicians had 6 days for baseline contouring (C1, reflecting daily practice).
- Session 2 presented contouring guidelines for EBRT and BT based on the EMBRACE (An intErnational study on MRI guided BRachytherapy in locally Advanced CErvical cancer) protocol, reviewed baseline contours, and included a question-and-answer session. Recommendations from the Gynaecological GEC-ESTRO working group, EMBRACE protocol, a pelvic nodal atlas and two consensus atlases for pelvic normal tissue were sent to clinicians to aid delineation [19,22–25]. They had 2 weeks to modify contours for the same image sets (guideline contouring: C2).
- In session 3 CHM reviewed baseline and guideline contours and held a question-and-answer session.

Lastly, clinicians performed final contouring (C3) for EBRT and BT 1.5–2 months after session 3, to evaluate the long term teaching impact.

Clinical case

A forty-five year old patient with a FIGO IIIB squamous cell CC was studied. Gynaecological exam: large growth (85x50x60 mm) involving the vagina (all fornices 1 cm, anterior vaginal wall 4 cm). The right parametrium had proximal infiltration, the left one until pelvic side wall. Bladder mucosa was not involved. Abdominopelvic CT showed CC with vaginal involvement, enlarged external, internal, lower common iliac, and pre-sacral nodes. No paraaortic nodes. The response to EBRT and concomitant chemotherapy was good: tumour dimensions of 55x40x30 mm, free right parametrium, induration of half of the left parametrium, and involvement of 1 cm of the anterior vaginal wall at the time of BT.

- Volumes required for contouring exercises (at least specified slices for OAR and whole ROI for TV):
  - EBRT:
    - OAR: Bladder, rectum, bowel, sigmoid.
    - GTV-P (gross tumour volume-P): Cervix, parametria and vaginal gross disease.
    - CTV-nodes: Nodal elective volume.
    - GTV node: Radiologically pathological lymph nodes (to boost).
    - CTV-P: GTV-P, uterus and vagina ($\geq$20 mm below GTV-P).
  - BT:
    - OAR: Bladder, rectum, sigmoid.
    - GTV: Macroscopic tumour at BT.
    - HR-CTV: Macroscopic tumour at BT + whole cervix + presumed extra-cervical tumour extension.
    - IR-CTV (intermediate risk CTV): HR CTV + GTV at diagnosis + $\geq$10 mm margin to residual disease at time of brachytherapy towards potential spread.

Contour evaluation methodology

Intraobserver variability was evaluated between C2 vs. C1, C3 vs. C2 and C3 vs. C1, for EBRT and BT treatments, quantitatively and qualitatively.

Interobserver variability was determined quantitatively by analyses centred on regions of interest (ROI) and on years of experience, and for BT also between centres that used MRI-based IGBT and others.

Contours were quantitatively classified by DICE scores [$DICE = 2 \times (Volume_{expert} \cap Volume_{participant})/(Volume_{expert} + Volume_{participant})$] given by FALCON EduCase™ Output [26]:

DICE references for TV [27,28]:

A: Optimal: >0.81
B: Average: 0.65–0.81
C: Suboptimal: <0.65

DICE references for OAR [29]:

A: Optimal: >0.81
B: Suboptimal: $\leq$0.81

In MRI-based brachytherapy for cervical cancer, Dimopoulos et al. defined a range of 0.5–0.7 using the conformity index for target volumes, which when converted to DICE is roughly 62.5–0.81

Table 1
Preworkshop questionnaire reflecting daily practice in each centre.

Centre	Centre Type	# Pat. RCT/Yr.	Type C. Ca. Pat. Treated	BT Applicator	Do Interstitial BT	Dose Rate	BT Imaging Type	Plan To Start 3D IGBT?	BT Prescription	Dose TV	# Fract.	Constraints
Centre 1*	Academic	<50	All pat. S.I–IVa, M0	Ovoids; Interstitial N.	YES	PDR	CT; MRI; US	NA	HR-CTV	80–84 Gy	2	Bladder Rectum Sigmoid
Centre 2*	Public	50–100	All pat. S.I–IVa, M0	Ovoids	NO	HDR	CBCT	In 3y	PointA	65–74 Gy	4	Bladder Rectum
Centre 3*	Academic; Public	>100	All pat. S.I–IVa, M0	Tandem/ring; Ovoids; Tandem/cylinder	NO	HDR	X-ray	NO	PointA	65–74 Gy	4	Bladder Rectum
Centre 4*	Academic	<50	All pat. S.I–IVa, M0	Ovoids	NO	PDR	CT; MRI	NA	PointA; HR-CTV	65–74 Gy	2	Bladder Rectum Sigmoid
Centre 5*	Academic; Public	>100	All pat. S.I–IVa, M0	Ovoids	NO	HDR	CT; MRI	NA	HR-CTV	75–79 Gy	2	Bladder Rectum Sigmoid
Centre 6*	Academic	<50	Pos. Pelv./PA LN	Ovoids	YES	PDR	CT; MRI	NA	HR-CTV	80–84 Gy	2	Bladder Rectum Sigmoid
Centre 7*	Private	50–100	All pat. S.I–IVa, M0	Ovoids; Interstitial N.	YES	PDR; LDR	CT; MRI; US	NA	HR-CTV	<65 Gy; 65–74 Gy	-	Bladder Rectum Sigmoid
Centre 8*	Academic	<50	Pos. Pelv./PA LN; Neg.LN S. > IIB	Ovoids; Mould; Interstitial N.	YES	PDR	CT; MRI	NA	HR-CTV	<65 Gy	2	Bladder Rectum Sigmoid
Centre 9*	Public	<50	Pos. Pelv./PA LN; Neg.LN S. > IIB	Tandem/ring; Mould	NO	PDR	CT; US	NA	HR-CTV	80–84 Gy	2	Bladder Rectum Sigmoid
Centre 10*	Academic; Public	<50	Pos. Pelv./PA LN; Neg.LN S. > IIB	Tandem/ring; Ovoids	NO	PDR	CT; US	NA	PointA; HR-CTV; IR-CTV	80–84 Gy	2	Bladder Rectum Sigmoid
Centre 11	Public	<50	Pos. PA LN; Neg.LN S. > IIB	Mould	NO	LDR	CT; MRI	NA	IR-CTV	65–74 Gy	2	Bladder Rectum
Centre 12*	Public	<50	All pat. S.I–IVa, M0	Ovoids	NO	PDR	CT	NA	HR-CTV	65–74 Gy	2	Bladder Rectum Sigmoid
Centre 13*	Academic	50–100	Pos. Pelv./PA LN; Neg.LN S. > IIB	Ovoids	NO	PDR	CT; MRI	NA	HR-CTV; IR-CTV	<65 Gy; 80–84 Gy	2	Bladder Rectum Sigmoid Vagina
Centre 14*	Academic	>100	All pat. S.I–IVa, M0	Ovoids; Mould; Interstitial N.	YES	PDR	MRI	NA	HR-CTV	>85 Gy	2	Bladder Rectum Sigmoid
Centre 15	Public	<50	Pos. Pelv./PA LN	Ovoids	YES	HDR	MRI	NA	HR-CTV	>85 Gy	3	Bladder Rectum Sigmoid
Centre 16	Academic	50–100	All pat. S.I–IVa, M0	Ovoids	YES	PDR	MRI	NA	HR-CTV	75–79 Gy; 80–84 Gy	2	Bladder Rectum Sigmoid
Centre 17*	Academic	>100	All pat. S.I–IVa, M0	Ovoids; Tandem/cylinder	NO	HDR	CT	NA	PointA; HR-CTV	80–84 Gy	2; 3	Bladder Rectum

The seventeen centres listed below answered this questionnaire, of which the 14 marked with an asterisk participated in the ODW (the order of the centres does not correspond with the order in the anonymous table).
French centres: Institut Curie*, Gustave Roussy*, Centre Georges François Leclerc*, Centre Léon Bérard*, Centre Alexis Vautrin*, CHU Anne de Bretagne*, Centre Jean Perrin*, Institut Bergonié*, Institut de Cancérologie de l'Ouest*; Centre Oscar Lambret*, Tenon Hospital, Institut Claudius Regaud.
The Netherlands: Academic Medical Centre (AMC)*, Antoni van Leewenhoek Hospital (NKI).
Moldova: Institute of Oncology (IOM)*.
Romania: Emergency County Hospital Oradea*.
Serbia: Institut za onkologiju Vojvodine (IOV)*.
Abbreviations: RCT: Radiochemotherapy; yr.: year; C.Ca.: Cervical cancer; Pat.: patients; BT: Brachytherapy; IGBT: Image guided brachytherapy; TV: Target Volume; Fract.: Fractions; S.: Stage; Pos.: Positive; Pelv.: Pelvic; PA: Para-aortic; LN: Lymph Nodes; N.: Needles; LDR: Low Dose Rate; PDR: Pulsed Dose Rate; HDR: High Dose Rate; CT: Computed Tomography scan; MRI: Magnetic Resonance Imaging; CBCT: Cone Beam CT scan; US: Ultrasound; NA: Not applicable; HR-CTV: High Risk CTV; IR-CTV: Intermediate Risk CTV; Gy: Gray.

Table 2
Participants enroled in the online delineation workshops which submitted contours for each contouring period.

	EBRT			BT		
	C1	C2	C3	C1	C2	C3
Submission of ≥ 1 contour	28	22	13	30	21	13
Submission of all contours (OAR and TV)	14	11	5	24	15	6
Submission of only TV	19	15	7	24	15	9
Total number of participants (n)			46			46
PARTICIPANT POPULATION WHICH SUBMITTED CONTOURS						
Experienced specialists (>5 years of post-residency experience)	13					
Less experienced specialists (≤5 years of post-residency experience)	8					
Senior residents (>2 years of experience)	7					
Junior residents (≤2 years of experience)	4					

The participant population which submitted contours, by level of experience.
Abbreviations: EBRT: External beam radiotherapy; BT: Brachytherapy; OAR: Organs at risk; TV: Target Volumes.

[27,30]. For OAR, Breunig et al. found an average DICE of 0.61 for volumes <8 cc and of 0.91 for volumes >8 cc, averaging at 0.76 [29]. To simplify the cutoffs and make the study easier to interpret, 0.65 and 0.81 were chosen for TV and 0.81 for OARs. Of note, all statistical analyses performed were independent of the thresholds that were only used to aid interpretation and to display the results.

For the objective qualitative intraobserver assessment, the EduCase™ contour error distance tool showed on axial slices where the participant contour was 3 mm larger or smaller than the expert contour, based on the scalar assessment in the transverse plane for HR-CTV by Petric et al., in 8 directions (anterior, posterior, right, left, anterolateral right and left and posterolateral right and left) to detect the most prevalent areas of uncertainties [13].

Qualitative Classification:

- "Correct": Participant contour ≤3 mm smaller/larger than the expert contour in a given direction.
- "Incorrect": Participant contour >3 mm smaller/larger than the expert contour in a given direction without a probable clinical impact.
- "Very incorrect": Participant contour >3 mm smaller/larger than the expert contour in a certain direction which for that particular ROI will have a probable clinical impact (worse coverage of TV/ higher dose to OAR).

As part of the outcome of e-learning courses depends on participant perception, an anonymous satisfaction questionnaire adapted from FALCON-ESTRO ODW was administered to clinicians (Appendix 3).

Statistical analyses

DICE scores have been transformed using the *logit* function, logit(x)=x/(1-x), so they asymptotically follow a Gaussian distribution [31,32].

To assess interobserver variability, a linear mixed model (**ModelINTER.PART)** was used, with the fixed effects *ROI*, *contouring period*, *experience*, their interactions, the linear and quadratic effects of the *slice* and their interactions with the *ROI* (for BT: the effect *imaging technique* and the interaction *ROI*imaging technique* was added), and the random effect of interparticipant variability, considered different for OAR and TV. To assess intraobserver variability (difference of DICE scores between contouring periods), a paired comparison by participant and ROI was performed using a linear model (**modelINTRA.PART**) with the fixed effect *participant*, *ROI* and their interaction (*participant*ROI*). The average DICE score by participant and ROI for each contouring period was assessed by a similar model (**modelSCORE.PART**). The significance of the fixed effects was computed using Fisher's test for all models. The proportion of pairs *participant*ROI* declared as performing better (or worse) from **modelINTRA.PART**, and their association with other covariates (experience, ROI type, institution or imaging technique) were assessed with Fisher's exact test. For the qualitative analysis, to compare the proportions of correct contours between different contouring periods, we used the test of McNemar [33]. The statistical analysis was performed using R software (R Core Team, 2016) and can be automatically reproduced using the scripts and data in Supplementary information.

Results

Participant population

Participants from 14 of 22 RAIDs centres submitted contours (Table 1).

Of the 46 enroled participants, nine submitted delineations for all contouring periods for EBRT and BT (Table 2). The description by level of experience of the participant population which submitted contours is in Table 2.

There is no significant relationship between the participants who dropped out after C1 with the initial DICE scores on C1 (whether they were low or high) nor with the years of experience or centre (Appendix 1: Tables 26–28; Appendix 2: Tables 29–31).

The adequacy of the cutoff points for this study was confirmed by the distribution of the pooled data (for EBRT and BT over all contouring attempts). The first quartile for OAR is 0.8 and for TV the first quartile is 0.6 and the third quartile is 0.86, which is mostly consistent with the chosen cutoff points (0.65 and 0.81).

All of the results of all models per contouring period are summarized in Table 3.

Results of ModelINTER.PART (interobserver variability)

All interactions were highly significant ($p < 0.001$), all effects had an impact on DICE scores (Table 2 in Appendix 1, 2). The model captures the quadratic relationship between DICE score and slice number (Fig. 1).

Pairwise comparisons for EBRT and BT between contouring periods by ROI are reported in Table 2 (details in Appendices 1–2, Table 5). For both EBRT and BT in C2 vs. C1 there was a significant improvement, mostly for TV, with no significant decrease. For C3 vs. C1 in EBRT and BT there was also a significant increase observed in certain TV with only a significant decrease for GTV node. However, in C3 vs. C2 for both image sets there was a significant decrease for 2 TV in EBRT and 2 OAR (no decrease for TV in BT), with a significant increase for CTV node.

For EBRT (Appendix 1, Table 6), regarding the experience effect, experienced specialists performed significantly better than junior

Table 3
Results for the interobserver and intraobserver quantitative and qualitative analyses. All results reported were statistically significant $p < 0.05$.

	EBRT			BT		
INTEROBSERVER QUANTITATIVE						
	C2 vs. C1	C3 vs. C1	C3 vs. C2	C2 vs. C1	C3 vs. C1	C3 vs. C2
Comparisons between contouring periods by ROI*	↑Bowel ↑CTV node ↑CTV-p ↑GTV node	 ↑CTV node ↓GTV node ↑GTV-p	↓Bowel ↑CTV node ↓CTV-p ↓GTV node	↑Bladder ↑GTV ↑HR-CTV ↑IR-CTV	 ↑HR-CTV ↑IR-CTV	↓Bladder
Comparisons between experience by ROI*	Sigmoid: Exp. Spec. > Junior Res. Less exp. Spec. > Junior Res. Less exp. Spec. > Senior Res. GTV node: Exp. Spec. < Less exp. Spec. Exp. Spec. < Senior Res. Less exp. Spec. > Junior Res. Senior Res. > Junior Res. GTV-p: Exp. Spec. < Junior Res. Junior Res. > Senior Res.			Sigmoid: Exp. Spec. > Senior Res.		
Comparison between imaging techniques*				HR-CTV: C. MRI-IGBT > Others		
INTRAOBSERVER QUANTITATIVE						
	C2 vs. C1	C3 vs. C1	C3 vs. C2	C2 vs. C1	C3 vs. C1	C3 vs. C2
Do participants improve between contouring periods?	Yes	Yes	Yes	Yes		
Is the improvement associated to ROI.type (TV/OAR)?	Yes			Yes		
Is the improvement associated to institution?	Yes					
INTRAOBSERVER QUALITATIVE (comparison of the % of correct contours between contouring periods)						
TV Posterolat. right	↑			↑	↓	↓
TV Anterolat. right		↓	↓			
TV Posterior			↓			
TV Posterolat. left			↓			
TV Right				↑		
OAR Posterolat. right					↑	

Abbreviations: EBRT: External beam radiotherapy; BT: Brachytherapy; C1: baseline contouring; C2: guideline contouring; C3: final contouring; exp.: experience; Spec.: Specialist; Res.: Resident; C. MRI-IGBT: centres using MRI-image guided BT; Posterolat. : Posterolateral; Anterolat. : Anterolateral.
↑: Improvement; ↓: Decrease.
* $p < 0.05$ after correction for multiple testing FDR.

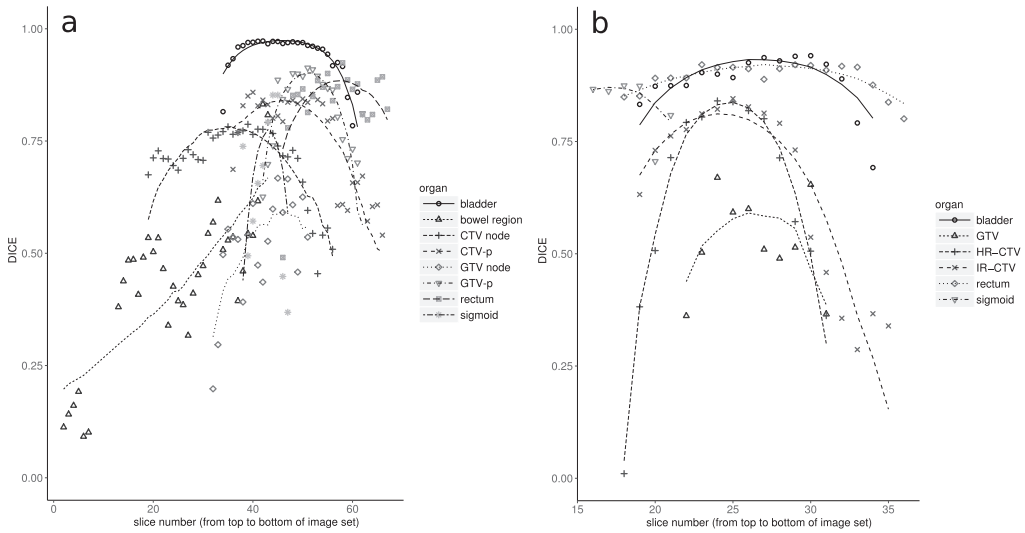


Fig. 1. Means of DICE scores (1 indicating perfect concordance between participant and expert; 0 indicating no concordance) by ROI according to slice number. The lines represent the mean of the DICE score predicted by the mixed model, capturing the parabolic effect of the slice number on the DICE scores. (a) Means of DICE scores by ROI according to slice number for EBRT. (b) Means of DICE scores by ROI according to slice number for BT.

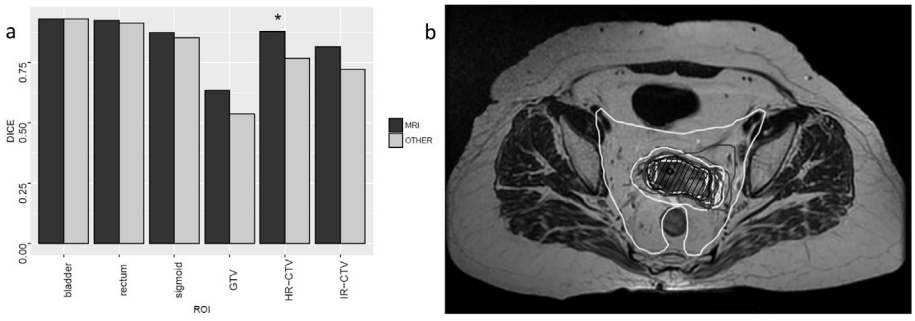


Fig. 2. (a) The interaction *ROI*imaging technique* in the BT treatment for centres using MRI-IGBT (black) and those not using it (light grey) (**ModelINTER.PART**). (b) Baseline contours (C1) of GTV for centres using MRI-IGBT (dark grey) and those not using it (white). Expert contour in black, with diagonal lines.

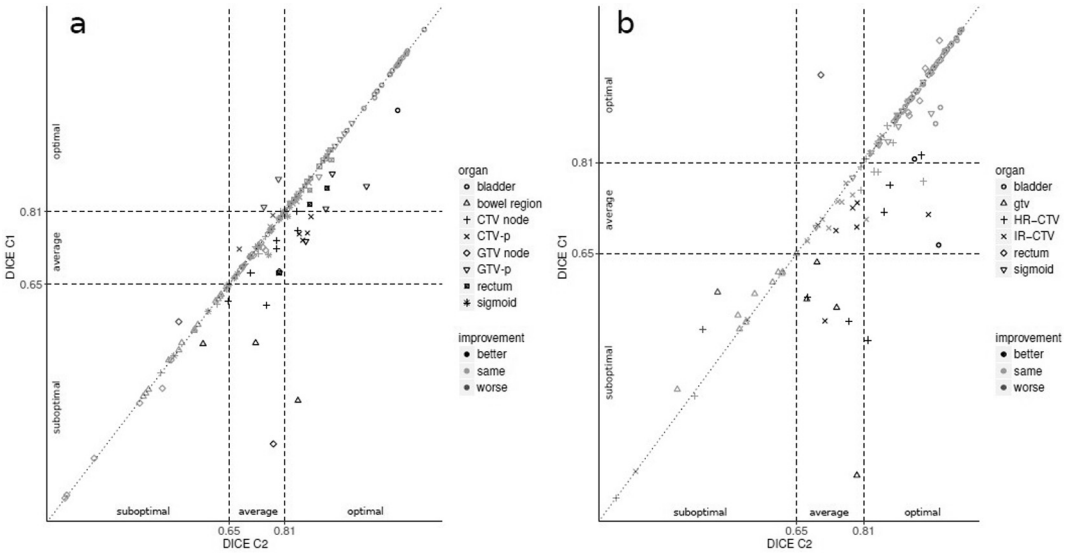


Fig. 3. The average score for each participant and each ROI for the C2 (*x*-axis) and C1 (*y*-axis) for (a) EBRT and (b) BT estimated from the **modelSCORE.PART** and whether the delineation improved or not (the difference is significantly better, equal or worse, from **modelINTRA.PART**, intraobserver variability). Examples (Ex.): Ex. 1.: This participant's DICE index did not vary significantly (same) between C2 vs. C1 for GTV, staying within the suboptimal category. Ex. 2.: This participant's DICE index was in significant detriment (worse) between C2 vs. C1 for rectum, changing from the optimal to the average category. Ex. 3.: This participant's DICE index improved significantly (better) between C2 vs. C1 for IR-CTV, changing from the average to the optimal category.

residents for sigmoid and significantly worse than less experienced specialists and senior residents for GTV node and than junior residents for GTV-p. Less experienced specialists did significantly better than junior residents for GTV node and sigmoid, and than senior residents for sigmoid. Between senior and junior residents there were only significant differences for GTV node and GTV-p. For BT the only significant difference was that experienced specialists performed better than senior residents for sigmoid (Table 3; Appendix 2, Table 6).

Regarding the imaging technique, centres that used MRI based IGBT did significantly better than those which used other techniques (CT, X-ray, US) for HR-CTV (Fig. 2, Table 3).

The ICC (intraclass correlation) for interobserver variability was excellent for OAR in BT (0.92; 95% CI: 0.86–0.96), OAR in EBRT (0.96; 95% CI: 0.93–0.98) and TV in EBRT (0.78–95% CI: 0.68–0.88) while it was fair for TV in BT (0.51; 95% CI: 0.39–0.68) (Table 4 in Appendix 1, 2). The low ICC for TV in BT highlights the difficulty of participants to agree on contours, whether they usually contour on MRI or not (the imaging technique was taken into account in modelINTER.PART).

Results of modelINTRA.PART and modelSCORE.PART (intraobserver variability)

Fig. 3 (Fig. 7, Appendixes 1 and 2) represents the average score for each participant and each ROI for C2 and C1 for EBRT and BT estimated from the **modelSCORE.PART** and whether the difference is significantly better, equal or is worse (from **modelINTRA.PART**). The Fisher's exact tests show that participants improved significantly between all contouring periods for EBRT (Appendix 1: Tables 8, 13 and 18). For BT, participants improved significantly between C2 vs. C1 (Appendix 2, Table 8). For EBRT, the improvements were significantly associated to ROI.type (TV vs. OAR) and institutions (C2 vs. C1). For BT, the improvements are significantly associated to ROI.type between C2 vs. C1 (Table 3). Interestingly, the number of participants who performed worse between different contouring periods was never significant (Fisher's exact test).

Results of qualitative data (intraobserver variability)

For EBRT, the percentage of "correct" contours was only significantly better between C2 vs. C1 for posterolateral right in TV. It was significantly worse in TV for anterolateral right for C3 vs. C2 and in three directions for C3 vs. C2 (Table 3; Appendix 1, Fig. 15).

For BT, the percentage of "correct" contours was significantly better between C2 vs. C1 for posterolateral right and right in TV and between C3 vs. C1 for posterolateral right in OAR. It only was significantly worse between C3 vs. C1 and C2 for posterolateral right in TV (Table 3; Appendix 2, Fig. 15).

Results of the satisfaction questionnaire

The scores over the 20 Organization and Content items for the 20 participants who responded of the 32 that submitted contours, on a scale of 1–5, 5 being excellent, range from 3.95 to 4.60 with an average of 4.358 (Table 3, Appendix 3). When asked whether they would attend another online workshop, 80% of participants answered affirmatively, and 85% would recommend one.

Discussion

For the first time this recent modality of ODW has been used for assessment of contouring skills as a dummy run within a multicentre trial [21]. Recently, other trials have used ODW within their quality assurance programmes, such as HYPO-G-01 [34]. Other authors, like Fokas et al., advocated training programmes within radiotherapy quality assurance protocols [35]. Our results have shown the ODW feasibility and capability in identifying centres that manifest baseline and subsequent average to optimal contours and are ready to include patients, while offering an effective educational tool for others. The added value of this study is that it reports the participants' point of view, which in light of the post-ODW satisfaction questionnaire results is extremely favourable.

Petric et al. have described graphically how the largest uncertainties in contouring are on the cranial and caudal slices of a volume, which coincides with our results (Fig. 1; Fig. 5 Appendixes 1 and 2) [15]. The bowel follows a particular pattern since the expert contour consisted of individual bowel loops and most participants contoured a bowel bag, but both contouring techniques are valid [36].

An interesting aspect of this study is that the evaluation of interobserver variability allowed assessment of overall improvement/detriment of the participants in the workshop group between them, and not only individual variability versus the expert contour (which affects the comparison of ROI, and has certain flaws) [16,37]. As could be expected, for interobserver comparisons in both EBRT and BT, there was overall more improvement between C2 and C1 than between C3 and C1, and the worse results were mostly between C3 and C2. This suggests that participants gained contouring skills after presentation of the guidelines, and retained part of this knowledge 1.5–2 months later. However, from the intraobserver point of view, only improvements were significant between contouring periods.

When considering interobserver variability with respect to experience in EBRT, experienced specialists did significantly worse than less experienced specialists and senior residents for GTV node. This finding is to be interpreted with caution, as there was a borderline significant paraaortic lymph node (though the clinical case states: 'no positive paraaortic lymph nodes') and a suspicious lymph node in the left groin, deemed as inflammatory by CHM during live sessions. Thus this may simply highlight that less experienced specialists and senior residents were more focused on the clinical information provided. Logically, less experienced specialists, with more experience, did better than junior residents for GTV node and than junior and senior residents for sigmoid, as senior residents did better than junior residents for GTV node (all significant). Surprisingly, junior residents did significantly better than senior residents for GTV-p. For BT the only significant difference was experienced specialists which contoured the sigmoid better than senior residents.

Concerning MRI guided IGBT, MRI allows better visualization of the vagina and uterus than of the rectum or bladder [38]. This may explain our results of interobserver improvement in HR-CTV and IR-CTV for C2 and C3 vs. C1, as opposed to a detriment for the bladder (C3 vs. C2). It is also interesting to note the significant improvement for contouring of HR-CTV for centres doing MRI-IGBT, showing the impact of specific training in MRI-based contouring.

Qualitatively, Petric et al. did not find significant interobserver differences along the 8 directions of space for HR-CTV contours [13]. Our intraobserver differences were significant for certain directions (better or worse) for EBRT, with no obvious explanation as there was no clear clinical impact due to these differences. However, for BT, the significant improvement towards the right and posterolateral right for TV in C2 vs. C1 most probably is because the left parametrial invasion made participants focus more on the left portion of the TV than on the right during C1, and they improved after the guideline session. But they went back to their old ways in C3, doing significantly worse for posterolateral right TV in C3 vs. C1 and C2.

Considering the e-learning educational experience, this ODW allows a self-directed path, each clinician may attend live online sessions, within a blended learning model (with support and interaction with tutors) or follow recordings. This flexibility adapted to the physicians' heavy workload [3]. But this was not always effective, 14 of the initial 46 enroled participants did not submit contours.

Initial limitations of this study were organizational: difficulties to locate radiotherapy professionals, as the ODW was performed before opening RAIDs in clinical centres. Further, the first ODW was held in June–July. Many clinicians could not participate, or only could attend some sessions, with fewer contour sets submitted during July (C2), and August–September (C3). Thus, only 13 contours were submitted for C3, limiting statistical significance and with a less representative population. Another limitation was the use of the DICE index, the only contouring conformity index available as FALCON EduCase™ output at the time. It is less reliable in small ROI volumes, such as GTV node, GTV or sigmoid in our study, showing lower concordance simply because slight delineation discrepancies have more impact on the score. Conversely, in very large ROI volumes, as CTV node or bowel, it seems to lack the sensitivity to identify divergences from the reference contour (Fig. 1) [29,39]. This is due to the duplication of the overlapping volume, which may inaccurately show considerable agreement in these large ROI. The strongpoint of this index is the simplicity of calculation, it is the most used in automatic segmentation studies [40]. It may also be converted into other concordance indexes using certain ratios [30].

In conclusion, ODW provide feasible and convenient means for initial assessment of contouring practices in geographically dispersed centres, as well as additional training in contouring within the setting of quality control for a multicentre trial. Future studies should focus on improving this training, and developing the optimal sequence of further training for centres which need more improvement (further online training, or combined with specific onsite programmes).

Conflict of interest statement

The first author of this manuscript has received a DUERTECC/EURONCO grant.

There are no other financial and personal relationships to disclose, on behalf of all authors.

Acknowledgements

"This project has received funding from the European Union's Seventh Programme for research, technological development and demonstration under grant agreement No 304810".

"Acting the recipient for the grant for DUERTECC/EURONCO (Diplôme Universitaire Européen de Recherche Translationnelle Et Clinique en Cancérologie)".

The RAIDs consortium, and particularly the participants of the ODW from the RAIDs centres:: Institut Curie, Gustave Roussy, Centre Georges François Leclerc, Centre Léon Bérard, Centre Alexis Vautrin, CHU Anne de Bretagne, Centre Jean Perrin, Institut Bergonié, Institut de Cancérologie de l'Ouest; Centre Oscar Lambret, Academic Medical Centre (AMC), Institute of Oncology, Moldova (IOM), Emergency County Hospital Oradea and the Institut za onkologiju Vojvodine (IOV).

ESTRO for granting us access to FALCON EduCase™ and technical support (Miika Palmu, Christine Verfaille) and Scott Kaylor and Arthur Boyer from EduCase™ (RadOnc eLearning Center, Inc. Fremont, CA, USA) for collaborating by incorporating a development within EduCase™.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.radonc.2017.05.008.

References

[1] Seddon B, Bidmead M, Wilson J, Khoo V, Dearnaley D. Target volume definition in conformal radiotherapy for prostate cancer: quality assurance in the MRC RT-01 trial. Radiother Oncol 2000;56:73–83. http://dx.doi.org/10.1016/S0167-8140(00)00191-2.

[2] McEnery KW, Roth SM, Kelley LK, Hirsch KR, Menton DN, Kelly EA. A method for interactive medical instruction utilizing the World Wide Web. Proc Symp Comput Appl Med Care 1995:502–7.

[3] Booth A, Carroll C, Papaioannou D, Sutton A, Wong R. Applying findings from a systematic review of workplace-based e-learning: Implications for health information professionals: review article. Health Info Libr J 2009;26:4–21. http://dx.doi.org/10.1111/j.1471-1842.2008.00834.x.

[4] Kronz JD, Silberman MA, Allsbrook WC, Epstein JI. A web-based tutorial improves practicing pathologists' Gleason grading of images of prostate carcinoma specimens obtained by needle biopsy: validation of a new medical education paradigm. Cancer 2000;89:1818–23.

[5] Ridgway PF, Sheikh A, Sweeney KJ, Evoy D, McDermott E, Felle P, et al. Surgical e-learning: validation of multimedia web-based lectures. Med Educ 2007;41:168–72. http://dx.doi.org/10.1111/j.1365-2929.2006.02669.x.

[6] Foroudi F, Pham D, Bressel M, Tongs D, Rolfo A, Styles C, et al. Learning methods in radiation oncology The utility of e-Learning to support training for a multicentre bladder online adaptive radiotherapy trial (TROG 10.01-BOLART). Radiother Oncol 2013;109:165–9. http://dx.doi.org/10.1016/j.radonc.2012.10.019.

[7] Pham D, Hardcastle N, Foroudi F, Kron T, Bressel M, Hilder B, et al. A multidisciplinary evaluation of a web-based elearning training programme for SAFRON II (TROG 13.01): a multicentre randomised study of stereotactic radiotherapy for lung metastases. Clin Oncol 2016. http://dx.doi.org/10.1016/j.clon.2016.03.005.

[8] Peters LJ, Giralt J, Fitzgerald TJ, Trotti A, Bernier J, Bourhis J, et al. Critical Impact of Radiotherapy Protocol Compliance and Quality in the Treatment of Advanced Head and Neck Cancer: Results From TROG 02.02. J Clin Oncol n. d.; 28:2996–3001. doi:http://dx.doi.org/10.1200/JCO.2009.27.4498.

[9] Weber DC, Tomsej M, Melidis C, Hurkmans CW. QA makes a clinical trial stronger: evidence-based medicine in radiation therapy. Radiother Oncol 2012;105:4–8. http://dx.doi.org/10.1016/j.radonc.2012.08.008.

[10] Abrams RA, Winter KA, Regine WF, Safran H, Hoffman JP, Lustig R, et al. Failure to adhere to protocol specified radiation therapy guidelines was associated with decreased survival in RTOG 9704-a phase iii trial of adjuvant chemotherapy and chemoradiotherapy for patients with resected adenocarcinoma of the pancreas. Int J Radiat Oncol Biol Phys 2012;1:809–16. http://dx.doi.org/10.1016/j.ijrobp.2010.11.039.

[11] Ohri N, Shen X, Dicker AP, Doyle LA, Harrison AS, Showalter TN. Radiotherapy protocol deviations and clinical outcomes: a meta-analysis of cooperative group clinical trials. J Natl Cancer Inst 2013;105:387–93. http://dx.doi.org/10.1093/jnci/djt001.

[12] Rasch C, Barillot I, Remeijer P, Touw A, Van Herk M, Lebesque JV. Definition of the prostate in CT and MRI: a multi-observer study. Int J Radiat Oncol Biol Phys 1999;43:57–66. http://dx.doi.org/10.1016/S0360-3016(98)00351-4.

[13] Petric P, Dimopoulos J, Kirisits C, Berger D, Hudej R, Pötter R. Inter- and intraobserver variation in HR-CTV contouring: Intercomparison of transverse and paratransverse image orientation in 3D-MRI assisted cervix cancer brachytherapy. Radiother Oncol 2008;89:164–71. http://dx.doi.org/10.1016/j.radonc.2008.07.030.

[14] van Mourik AM, Elkhuizen PHM, Minkema D, Duppen JC, van Vliet-Vroegindeweij C. Multiinstitutional study on target volume delineation variation in breast radiotherapy in the presence of guidelines. Radiother Oncol 2010;94:286–91. http://dx.doi.org/10.1016/j.radonc.2010.01.009.

[15] Petrič P, Hudej R, Rogelj P, Blas M, Tanderup K, Fidarova E, et al. Uncertainties of target volume delineation in MRI guided adaptive brachytherapy of cervix cancer: a multi-institutional study. Radiother Oncol 2013;107:6–12. http://dx.doi.org/10.1016/j.radonc.2013.01.014.

[16] Vinod SK, Jameson MG, Min M, Holloway LC. Uncertainties in volume delineation in radiation oncology: A systematic review and recommendations for future studies. Radiother Oncol 2016;121:169–79. http://dx.doi.org/10.1016/j.radonc.2016.09.009.

[17] Vinod SK, Min M, Jameson MG, Holloway LC. A review of interventions to reduce inter-observer variability in volume delineation in radiation oncology. J Med Imaging Radiat Oncol 2016;60:393–406. http://dx.doi.org/10.1111/1754-9485.12462.

[18] Pötter R, Georg P, Dimopoulos JCA, Grimm M, Berger D, Nesvacil N, et al. Clinical outcome of protocol based image (MRI) guided adaptive brachytherapy combined with 3D conformal radiotherapy with or without chemotherapy in patients with locally advanced cervical cancer. Radiother Oncol 2011;100:116–23. http://dx.doi.org/10.1016/j.radonc.2011.07.012.

[19] Haie-Meder C, Pötter R, Van Limbergen E, Briot E, De Brabandere M, Dimopoulos J, et al. Recommendations from Gynaecological (GYN) GEC-ESTRO Working Group (I): Concepts and terms in 3D image based 3D treatment planning in cervix cancer brachytherapy with emphasis on MRI assessment of GTV and CTV. Radiother Oncol 2005;74:235–45. http://dx.doi.org/10.1016/j.radonc.2004.12.015.
[20] Rational molecular Assessments and Innovative Drug Selection: RAIDs. n.d. http://www.raids-fp7.eu/ [accessed December 9, 2016].
[21] Grau Eriksen J, Salembier C, Rivera S, De Bari B, Berger D, Mantello G, et al. ESTRO FALCON e-learning Four years with FALCON – an ESTRO educational project: achievements and perspectives on behalf of ESTRO. Radiother Oncol 2014;112:145–9. http://dx.doi.org/10.1016/j.radonc.2014.06.017.
[22] Taylor A, Rockall AG, Reznek RH, Powell MEB. Mapping pelvic lymph nodes: Guidelines for delineation in intensity-modulated radiotherapy. Int J Radiat Oncol Biol Phys 2005;63:1604–12. http://dx.doi.org/10.1016/j.ijrobp.2005.05.062.
[23] Gay HA, Barthold HJ, O'Meara E, Bosch WR, El Naqa I, Al-Lozi R, et al. Pelvic normal tissue contouring guidelines for radiation therapy: a radiation therapy oncology group consensus panel atlas. Int J Radiat Oncol Biol Phys 2012;83. http://dx.doi.org/10.1016/j.ijrobp.2012.01.023.
[24] Gay HA, Barthold HJ, O'Meara E et al. Female Pelvis Normal Tissue RTOG Consensus Contouring Guidelines. n.d. https://www.rtog.org/LinkClick.aspx?fileticket=P5eAjYB90Ow%3D&tabid=355 [accessed December 9, 2016].
[25] An IntErnational study on MRI-guided BRAchytherapy in locally advanced CErvical cancer: EMBRACE. n.d. https://www.embracestudy.dk/.
[26] Dice LR. Measures of the amount of ecologic association between species. Ecology 1945;26:297–302. http://dx.doi.org/10.2307/1932409.
[27] Dimopoulos JCA, De Vos V, Berger D, Petric P, Dumas I, Kirisits C, et al. Inter-observer comparison of target delineation for MRI-assisted cervical cancer brachytherapy: application of the GYN GEC-ESTRO recommendations. Radiother Oncol 2009;91:166–72. http://dx.doi.org/10.1016/j.radonc.2008.10.023.
[28] Petersen RP, Truong PT, Kader HA, Berthelet E, Lee JC, Hilts ML, et al. Target volume delineation for partial breast radiotherapy planning: clinical characteristics associated with low interobserver concordance. Int J Radiat Oncol Biol Phys 2007;69:41–8. http://dx.doi.org/10.1016/j.ijrobp.2007.01.070.
[29] Breunig J, Hernandez S, Lin J, Alsager S, Dumstorf C, Price J, et al. A system for continual quality improvement of normal tissue delineation for radiation therapy treatment planning. Int J Radiat Oncol Biol Phys 2012;83:e703–8. http://dx.doi.org/10.1016/j.ijrobp.2012.02.003.
[30] Fotina I, Lütgendorf-Caucig C, Stock M, Pötter R, Georg D. Critical discussion of evaluation parameters for inter-observer variability in target definition for radiation therapy. Strahlenther Onkol 2012;188:160–7. http://dx.doi.org/10.1007/s00066-011-0027-6.
[31] Agresti A. Categorical Data Analysis. vol. 359. 2002. doi:http://dx.doi.org/10.1002/0471249688.
[32] Brock KK. Image processing in radiation therapy. CRC Press; 2013.
[33] McNemar Q. Note on the sampling error of the difference between correlated proportions or percentages. Psychometrika 1947;12:153–7. http://dx.doi.org/10.1007/BF02295996.
[34] HYPOG-01 n.d. https://www.gustaveroussy.fr/en/node/3646 [accessed January 8, 2017].
[35] Fokas E, Clifford C, Spezi E, Joseph G, Branagan J, Hurt C, et al. Comparison of investigator-delineated gross tumor volumes and quality assurance in pancreatic cancer: analysis of the pretrial benchmark case for the SCALOP trial. Radiother Oncol 2015;117:432–7. http://dx.doi.org/10.1016/j.radonc.2015.08.026.
[36] Banerjee R, Chakraborty S, Nygren I, Sinha R. Small bowel dose parameters predicting grade $\geq$3 acute toxicity in rectal cancer patients treated with neoadjuvant chemoradiation: An independent validation study comparing peritoneal space versus small bowel loop contouring techniques. Int J Radiat Oncol Biol Phys 2013;85:1225–31. http://dx.doi.org/10.1016/j.ijrobp.2012.09.036.
[37] Vinod SK, Lim K, Bell L, Veera J, Ohanessian L, Juresic E, et al. High-risk CTV delineation for cervix brachytherapy: Application of GEC-ESTRO guidelines in Australia and New Zealand. J Med Imaging Radiat Oncol 2016:1–8. http://dx.doi.org/10.1111/1754-9485.12509.
[38] Dimopoulos JCA, Schard G, Berger D, Lang S, Goldner G, Helbich T, et al. Systematic evaluation of MRI findings in different stages of treatment of cervical cancer: potential of MRI on delineation of target, pathoanatomic structures, and organs at risk. Int J Radiat Oncol Biol Phys 2006;64:1380–8. http://dx.doi.org/10.1016/j.ijrobp.2005.10.017.
[39] Esthappan J, Ma DJ, Narra VR, Raptis CA, Grigsby PW. Comparison of apparent diffusion coefficient maps to T2-weighted images for target delineation in cervix cancer brachytherapy. J Contemp Brachytherapy 2011;3:193–8. http://dx.doi.org/10.5114/jcb.2011.26470.
[40] Hoang Duc AK, Eminowicz G, Mendes R, Wong S-L, McClelland J, Modat M, et al. Validation of clinical acceptability of an atlas-based segmentation algorithm for the delineation of organs at risk in head and neck cancer. Med Phys 2015;42:5027–34. http://dx.doi.org/10.1118/1.4927567.

Annexes

8.2 Satisfaction questionnaire

Techical partnership with:

ONLINE CONTOURING WORKSHOP ON ADVANCED STAGE CERVICAL CANCER

June 20th and 26th and July 10th, 2013

We hope that you have found this workshop useful, but since nothing is perfect, we need your evaluation to continue to develop this workshop to meet participants' needs. We therefore kindly ask you to fill this in and send it back to us. Thank you for your co-operation.

I BACKGROUND INFORMATION:

1. Gender: □ Male □ Female

I am working in .....(country)

2. Specialty:

I am a ( ) **Radiation oncologist** ( ) Specialist ( ) Trainee
( ) **Other**_____

Number of years working in the specialty_____

3. I learned about this workshop from:

( ) RAIDs project manager (M. Kamal) and/ or WP-4.2 leader IGR (C.Haie-Meder,E.Deutsch,E.Rivin)
( ) Department director
( ) Colleagues

4. I have previously attended the following ESTRO events:

( ) ESTRO teaching course(s)
( ) ESTRO meeting(s)
( ) ESTRO on-line course(s)
( ) ESTRO contouring workshop(s)
( ) ESO courses
( ) Other courses
( ) Other contouring workshop(s)

5. Do you perform contouring in cervical cancer?

❒ Never
❒ Sometimes – please specify:.....
❒ Yes, for External Beam Radiotherapy under supervision
❒ Yes, for External Beam Radiotherapy without supervision
❒ Yes, for Brachytherapy under supervision
❒ Yes, for Brachytherapy without supervision

6. Do you have experience with contouring in:

- ❒ CT-images only
- ❒ MRI images only
- ❒ Matched CT and MRI images
- ❒ PET-images
- ❒ Multimodality (PET-CT-MRI)
- ❒ Ultrasound images

II. ORGANISATION and EDUCATIONAL MATTERS

Grading scale: (5) = Excellent (4) = Good (3) = Average (2) = Sufficient (1) = Poor

7. The goals of this workshop were clearly stated at the beginning ( )

……

……

8. The supplementary material was ( )

……

……

9. The Webex platform for the live presentations was ( )

……

……

10. The Falcon-Educase contouring tool was ( )

……

……

11. The workload demands were realistic for this workshop ( )

……

……

12. The experts' presentation was ( )

……

……

13. The discussion by the experts and the tutors was ( )

……

……

14. The language used was clear and understandable and facilitated interaction ( )

……

……

15. The experts and tutors continuously encouraged communication ( )

……

……

16. The support from ESTRO and tutors for the delineation exercise was ( )

……

……

17. The learning experience was equivalent to the experience in face-to-face workshops ( )

……

……

18. The workshop met my expectations ( )

……

……

19. Other comments:

……

……

……

……

III. CONTENTS OF THE WORKSHOP

Grading scale: (5) = Excellent (4) = Good (3) = Average (2) = Sufficient (1) = Poor

General		***Please, circle the digit of your choice on the scales below.***
1.	Did the workshop provide the following goals and learning outcomes:	
	The workshop should allow the participants to obtain a thorough understanding of the contouring guidelines in cervical cancer.	1 2 3 4 5 no --------relatively------ yes
	The workshop should allow the participants to obtain a good knowledge of contouring in cervical cancer and useful imaging modalities to do so.	1 2 3 4 5 no --------relatively------ yes
	The workshop should allow the participants to improve their contouring skills	1 2 3 4 5 no --------relatively------ yes
2.	Please, provide your overall rating of the quality of the education offered at this workshop.	1 2 3 4 5 poor ----------excellent

Content		
3.	Was the information useful and relevant to your work and practice techniques?	1 2 3 4 5 no ---------- yes
4.	Do you feel that the presented information was well balanced and supported by adequate evidence?	1 2 3 4 5 no ---------- yes
5.	Did the programme allow adequate time for discussion and questions?	1 2 3 4 5 no ---------- yes
Organization		
6.	How would you rate the management and organization of this workshop?	1 2 3 4 5 poor ----------excellent

IV. GENERAL CONCLUSION:

What topics would you like to be added to the on-line contouring workshops?

……..……..……..

……..……..……..

……..……..……..

Any special comments:

……..……..……..

……..……..……..

……..……..……..

Would you be interested in attending another on-line contouring workshop?

……..……..……..

……..……..……..

……..……..……..

Would you recommend an on-line contouring workshop?

……..……..……..

……..……..……..

……..……..……..

8.3 Curriculum vitæ

8.3.1 Education

1999-2006 M.D. in Medicine and Surgery. University of Malaga; Malaga, Spain.

2008-2012 Specialty in Radiation Oncology. Reina Sofia University Hospital; Cordoba, Spain.

2008-2011 European Master's degree in Nutrition and Metabolism (Thesis: Nutritional Evaluation in Cancer Patients before and after RT). University of Cordoba; Córdoba, Spain.

2012-2013 European Diploma in Clinical and Translational Research in Oncology *D.U.E.R.T.E.C.C.* Gustave Roussy Cancer Campus – Paris Sud University; Paris, France.

2013 – European Doctorate in Biomedicine. (Thesis project: Radiotherapy Quality Control in Cervical Cancer.) University of Cordoba, Radiology and Oncology group.

8.3.2 Work experience

2008 2012 Reina Sofia University Hospital, Department of Radiation Oncology; Cordoba, Spain. Residency in Radiation Oncology.

- 2011(March-May) Rotation in Brachytherapy. Valencian Oncology Institute; Valencia, Spain.
- 2011(November-December) Observer rotation in SBRT (Cyberknife). Collaborator. University of California at San Francisco; San Francisco, United States.

2012 Elche General University Hospital. ERESA. Department of Radiation Oncology; Elche, Spain. Specialist in Radiation Oncology.

- Experience in every subspecialty with emphasis in breast and gynaecological cancer.

2013-2017 Gustave Roussy Cancer Campus Grand Paris. Department of Radiation Oncology – Group A; Villejuif, France. Specialist in Radiation Oncology.

- Subspecialties: Breast cancer. Standard treatment (classic and hypofractionated schedules) or accelerated partial breast irradiation (APBI) pre or postoperatively in clinical trials. Gastrointestinal cancer. Standard treatment. Liver and pancreatic SBRT.

2017 Tenon University Hospital. Department of Radiation Oncology. Specialist in Radiation Oncology.

- Subspecialties: Breast, lung, gastrointestinal and prostate cancer.

8.3.3 Clinical trials experience

2008-2011 Design of the protocol: *Nutritional evaluation before and after radiotherapy*, its implementation, statistics, and writing for my master's thesis.

2012 Quality assurance of radiochemotherapy in cervical cancer in a european trial. Writing of a randomised Ph. II-III trial: antiviral agent + RTCT in advanced cervical cancer.

2014 Co-investigator in my center in: PAPBI, SHARE, BONBIS, PRAVACUR and RAIDs.

2017 PI of the EORTC Ph. IV trial validating the ANL-27 QofL questionnaire.

2016-2017 Real time radiotherapy quality assurance for NBTXR3 in liver SBRT.

8.4 List of publications within the scope of this thesis

8.4.1 Peer review articles

- Maroun P, **Rivin E**, Dumas I, et al. Locally advanced cervical cancer in renal transplant patients: a dilemma between control and toxicity. Brachytherapy. 2014 Jan-Feb;13(1):88-93.

- Mazeron R, Kamsu Kom L, **Rivin del Campo E**, et al. Comparison between the ICRU rectal point and modern volumetric parameters in brachytherapy for locally advanced cervical cancer. Cancer Radiother. 2014 Jun;18(3):177-82.

- Mazeron R, Aguini N, **Rivin E**, et al. Improving safety in radiotherapy: The implementation of the Global Risk Analysis method. Radiother Oncol. 2014 Aug;112(2):205-11.

- Mazeron R, Castelnau-Marchand P, Dumas I, **del Campo ER**, et el. Impact of treatment time and dose escalation on local control in locally advanced cervical cancer treated by chemoradiation and image-guided pulsed-dose rate adaptive brachytherapy. Radiother Oncol. 2015 Feb;114(2):257-63.

- Mazeron R, Champoudry J, Gilmore J, Dumas I, Goulart J, Oberlander AS, **Rivin del Campo E**, et al. Intrafractional organs movement in three-dimensional image-guided adaptive pulsed-dose-rate cervical cancer brachytherapy: Assessment and dosimetric impact. Brachytherapy. 2015 Mar-Apr;14(2):260-6.

- Mazeron R, Dumas I, **Rivin E**, et al. D2cm3/DICRU ratio as a surrogate of bladder hotspots localizations during image-guided adapted adaptive brachytherapy for cervical cancer: Assessment and implications in late urinary morbidity analysis. Brachytherapy. 2015 Mar-Apr;14(2):300-7.

- Mazeron R, Aguini N, **Rivin del Campo E**, et al. Implementation of the global risk analysis in pulsed-dose rate brachytherapy: Methods and results. Cancer Radiother. 2015 Apr;19(2):89-97.

- Castelnau-Marchand P, Chargari C, Bouaita R, Dumas I, Farha G, Kamsu-Kom L, **Rivin del Campo E**, et al. What to expect from immediate salvage hysterectomy following concomitant chemoradiation and image-guided adaptive brachytherapy in locally advanced cervical cancer. Cancer Radiother. 2015 Dec;19(8):710-7.

- Mazeron R, Maroun P, Castelnau-Marchand P, Dumas I, **del Campo ER**, et al. Pulsed-dose rate image-guided adaptive brachytherapy in cervical cancer: Dose-volume effect relationships for the rectum and bladder. Radiother Oncol. 2015 Aug;116(2):226-32.

- Castelnau-Marchand P, Chargari C, Maroun P, Dumas I, **del Campo ER**, et al. Clinical outcomes of definitive chemoradiation followed by intracavitary pulsed-dose rate image-guided adaptive brachytherapy in locally advanced cervical cancer. Gynecol Oncol. 2015 Nov;139(2):288-94.

- Mazeron R, Petit C, **Rivin E**, et al. 45 or 50 Gy, Which is the Optimal Radiotherapy Pelvic Dose in Locally Advanced Cervical Cancer in the Perspective of Reaching Magnetic Resonance Image-guided Adaptive Brachytherapy Planning Aims? Clin Oncol (R Coll Radiol). 2016 Mar;28(3):171-7.

- Limkin EJ, Dumas I, **Rivin del Campo E**, et al. Vaginal dose assessment in image-guided brachytherapy for cervical cancer: Can we really rely on dose-point evaluation? Brachytherapy. 2016. Mar-Apr;15(2):169-76.

- Mazeron R, Castelnau-Marchand P, Escande A, **Rivin del Campo E**, et al. Tumor dose-volume response in image-guided adaptive brachytherapy for cervical cancer: A meta-regression analysis. Brachytherapy 2016 Sep-Oct;15(5):537-42.

- Mazeron R, Gouy S, Chargari C, **Rivin del Campo E**, et al. Post radiation hysterectomy in locally advanced cervical cancer: Outcomes and dosimetric impact. Radiother Oncol. 2016 Sep;120(3):460-466.

- Mazeron R, **Rivin del Campo E**, Haie-Meder C, Chargari C. In Regard to Swanick et al. Int J Radiat Oncol Biol Phys. 2017 Mar 1;97(3):638.

- Bacorro W, Dumas I, Levy A, **Rivin del Campo E**, et al. Contribution of image-guided adaptive brachytherapy to pelvic nodes treatment in locally advanced cervical cancer. Brachytherapy. 2017 Mar - Apr;16(2):366-372.

- **E Rivin del Campo**, S Rivera, M Martínez-Paredes et al. Assessment of the novel online delineation workshop dummy run approach using FALCON within a European multicentre trial in cervical cancer (RAIDs). Radiother Oncol. 2017 (in press).

8.5 Participation to conferences within the scope of this thesis

8.5.1 Oral communications

April 2013 IIIrd Congress of Young Researchers of the University of Cordoba. Advanced Cervical Cancer: Analysis of Target Volume Delineation. **E. Rivin**.

April 2013 10th Scientific and Medical Days of Institut Curie. RAIDs in gynaecology. M. Kamal and **E. Rivin**

April 2014 European Society for Radiotherapy & Oncology (ESTRO) 33 Congress. Quality input of an online delineation workshop in advanced stage cervical cancer. Initial results. **E. Rivin del Campo**, S. Rivera, M. Martínez-Paredes, et al.

8.5.2 Poster

November 2016 V Congress of Young Researchers of the University of Cordoba.

- Radiotherapy in advanced cervical cancer: initial results of a delineation workshop of organs at risk and target volumes within a European Multicentre Trial. **E. Rivin del Campo**, et al.

8.5.3 Co-author of oral communications

April 2015 3rd European Society for Radiotherapy & Oncology (ESTRO) Forum. Image-guided adaptive brachytherapy in cervical cancer: towards a personalization of planning aims. C. Chargari, R. Mazeron, I. Dumas, P. Castelnau-Marchand, **E. Rivin del Campo**, et al.

May 2017 European Society for Radiotherapy & Oncology (ESTRO) 36 Congress. Dose contribution to pelvic nodes of image guided adaptive brachytherapy in cervical cancer. W. Bacorro, I. Dumas, A. Levy, **E. Rivin del Campo**, et al.

June 2017 International Conference in Advances in Radiation Oncology ICARO 2. Nodal doses during image-guided adaptive brachytherapy for cervical cancer and implication to simultaneous integrated boost. W. Bacorro, I. Dumas, A. Levy, **E. Rivin del Campo**, et al.

8.5.4 Co-author of poster presentation

April 2014 European Society for Radiotherapy & Oncology (ESTRO) 33 Congress. Improving safety in radiotherapy: The implementation of the Global Risk Analysis method. R. Mazeron, N. Aguini, **E. Rivin del Campo**, et al. (Young Scientists Poster Session).

8.5.5 Co-auteur of posters

April 2014 European Society for Radiotherapy & Oncology (ESTRO) 33 Congress.

- Locally advanced cervical cancer in renal transplant patients: A dilemma between control and toxicity. P. Maroun, **E. Rivin**, I. Dumas, et al. (e- Poster).

September 2014 American Society for Radiation Oncology (ASTRO)

- Impact of Overall Treatment Time and Dose Escalation on Local Control in Locally Advanced Cervical Cancer Treated by Chemoradiation and Image Guided

Adaptive Brachytherapy. P. Castelnau-Marchand, R. Mazeron, I. Dumas, **E. Rivin del Campo**, et al. (e-Poster)

8.5.6 Presentation at working group

April 2013 Rational molecular Asessments and Innovative Drug selection in advanced stage cervical cancer (RAIDs) 6 month steering committee within the International. Charité-Mayo Conference. (Institut Curie-all RAIDs partners).

- Achieving standard chemoradiotherapy quality control. **E. Rivin**, C. Haie-Meder, E. Deutsch, S. Rivera.
- Phase II/III trial in cervical cancer associating the antiviral agent Vistide® and radiochemotherapy. **E. Rivin**, E. Deutsch,M. Mondini.

Bibliography

Agresti, A. **Categorical Data Analysis**. Wiley-Interscience, 2002.

Aleman, B. M. P., Girinsky, T., Van Der Maazen, R. W. M., Strijk, S., Meijnders, P., Bortolus, R., Olofsen-Van Acht, J. J., Lybeert, M. L. M., Lievens, Y., Eghbali, H., Noordijk, E. M., Tomšič, R., Meerwaldt, J. H., Poortmans, P. M. P., Smit, W. G. J. M., Pinna, A., Henry-Amar, M., and Raemaekers, J. M. M. **Quality control of involved-field radiotherapy in patients with advanced Hodgkin's lymphoma (EORTC 20884)**. *International Journal of Radiation Oncology Biology Physics* 63.4 (2005), pp. 1184–1190.

Altman, M. B., Kavanaugh, J. a., Wooten, H. O., Green, O. L., DeWees, T. a., Gay, H., Thorstad, W. L., Li, H., and Mutic, S. **A framework for automated contour quality assurance in radiation therapy including adaptive techniques**. *Physics in Medicine and Biology* 60.13 (2015), pp. 5199–5209.

Banerjee, R., Chakraborty, S., Nygren, I., and Sinha, R. **Small bowel dose parameters predicting grade ≥3 acute toxicity in rectal cancer patients treated with neoadjuvant chemoradiation: An independent validation study comparing peritoneal space versus small bowel loop contouring techniques**. *International Journal of Radiation Oncology Biology Physics* 85.5 (2013), pp. 1225–1231.

Barillot, I., Tavernier, E., Peignaux, K., Williaume, D., Nickers, P., Leblanc-Onfroy, M., and Lerouge, D. **Impact of post operative intensity modulated radiotherapy on acute gastro-intestinal toxicity for patients with endometrial cancer: Results of the phase II RTCMIENDOMETRE French multicentre trial**. *Radiotherapy and Oncology* 111 (2014), pp. 138–143.

Bates, D., Mächler, M., Bolker, B., and Walker, S. **Fitting Linear Mixed-Effects Models Using lme4**. *Journal of Statistical Software* 67.1 (2015), pp. 1–48.

Bekelman, J. E., Deye, J. A., Vikram, B., Bentzen, S. M., Bruner, D., Curran, W. J., Dignam, J., Efstathiou, J. A., Fitzgerald, T. J., Hurkmans, C., Ibbott, G. S., Lee, J. J., Merchant, T. E., Michalski, J., Palta, J. R., Simon, R., Ten Haken, R. K., Timmerman, R., Tunis, S., Coleman, C. N., and Purdy, J. **Redesigning radiotherapy quality assurance: Opportunities to develop an efficient, evidence-based system to support clinical**

trials - Report of the National Cancer Institute work group on radiotherapy quality assurance. *International Journal of Radiation Oncology Biology Physics* 83.3 (2012), pp. 782–790. arXiv: NIHMS150003.

Benjamin, Y. and Hochberg, Y. **Controlling the false discovery rate: a practical and powerful approach to multiple testing**. *J Royal Statistical Society, Series B (57)* 289 (1995).

Benoit, A. G., Krepart, G. V., and Lotocki, R. J. **Results of prior cytologic screening in patients with a diagnosis of Stage I carcinoma of the cervix**. *American Journal of Obstetrics and Gynecology* 148.5 (1984), pp. 690–694.

Booth, A., Carroll, C., Papaioannou, D., Sutton, A., and Wong, R. **Applying findings from a systematic review of workplace-based e-learning: Implications for health information professionals: Review Article**. *Health Information and Libraries Journal* 26.1 (2009), pp. 4–21.

Breunig, J., Hernandez, S., Lin, J., Alsager, S., Dumstorf, C., Price, J., Steber, J., Garza, R., Nagda, S., Melian, E., Emami, B., and Roeske, J. C. **A system for continual quality improvement of normal tissue delineation for radiation therapy treatment planning**. *International Journal of Radiation Oncology Biology Physics* 83.5 (2012), e703–e708.

Brock, K. K. . **Image Processing in Radiation Therapy**. CRC Press, 2013.

Choi, C., Koo Cho, C., Yul Yoo, S., Sook Kim, M., Yang, K. M., Jun Yoo, H., Seok Seo, Y., Kyu Kang, J., Han Lee, D., Hee Lee, K., Don Lee, E., Young Rhu, S., Chul Choi, S., Hong Kim, M., and Jong Kim, B. **Image-guided stereotactic body radiation therapy in patients with isolated para-aortic lymph node metastases from uterine cervical and corpus cancer**. *International Journal of Radiation Oncology Biology Physics* 74 (2009), pp. 147–153.

Chung, Y., Kim, J. W., Shin, K. H., Kim, S. S., Ahn, S.-J., Park, W., Lee, H.-S., Dong, #., Kim, W., Lee, K. C., Hyun, Y., Suh, S., Jin, Z., Kim, H., Shin, H. S., Kim, Y. B., and Suh, C.-O. **Dummy Run of Quality Assurance Program in a Phase 3 Randomized Trial Investigating the Role of Internal Mammary Lymph Node Irradiation in Breast Cancer Patients: Korean Radiation Oncology Group 08-06 Study**. *Radiation Oncology Biology* 91.2 (2015), pp. 419–426.

Clark, C. H., Miles, E. A., Guerrero Urbano, M. T., Bhide, S. A., Bidmead, A. M., Harrington, K. J., and Nutting, C. M. **Pre-trial quality assurance processes for an intensity-modulated radiation therapy (IMRT) trial: PARSPORT, a UK mul-**

ticentre Phase III trial comparing conventional radiotherapy and parotid-sparing IMRT for locally advanced head and neck cancer. *The British Journal of Radiology* 82 (2009), pp. 585–594.

ClinicalTrials.gov NCT01566240. **Induction Chemotherapy Plus Chemoradiation as First Line Treatment for Locally Advanced Cervical Cancer (INTERLACE)**. 2012.

Consortium RAIDs. **Raids, http://www.raids-fp7.eu/**. 2017.

De Azevedo, C. R. A. S., Thuler, L. C. S., de Mello, M. J. G., and Ferreira, C. G. **Neoadjuvant Chemotherapy Followed by Chemoradiation in Cervical Carcinoma**. *International Journal of Gynecological Cancer* 00.00 (2016), p. 1.

Dimopoulos, J. C. A., Schard, G., Berger, D., Lang, S., Goldner, G., Helbich, T., and Pötter, R. **Systematic evaluation of MRI findings in different stages of treatment of cervical cancer: Potential of MRI on delineation of target, pathoanatomic structures, and organs at risk**. *International Journal of Radiation Oncology Biology Physics* 64.5 (2006), pp. 1380–1388.

Dimopoulos, J. C. A., Vos, V. D., Berger, D., Petric, P., Dumas, I., Kirisits, C., Shenfield, C. B., Haie-Meder, C., and Pötter, R. **Inter-observer comparison of target delineation for MRI-assisted cervical cancer brachytherapy: Application of the GYN GEC-ESTRO recommendations**. *Radiotherapy and Oncology* 91.2 (2009), pp. 166–172.

Dueñas-González, A., Zarbá, J. J., Patel, F., Alcedo, J. C., Beslija, S., Casanova, L., Pattaranutaporn, P., Hameed, S., Blair, J. M., Barraclough, H., Orlando, M., Lilly, E., by Eli Lilly, and Lilly Interamerica, E. **Phase III, Open-Label, Randomized Study Comparing Concurrent Gemcitabine Plus Cisplatin and Radiation Followed by Adjuvant Gemcitabine and Cisplatin Versus Concurrent Cisplatin and Radiation in Patients With Stage IIB to IVA Carcinoma of the Cervix**. *J Clin Oncol* 29 (2011), pp. 1678–1685.

Eifel, P. J., Winter, K., Morris, M., Levenback, C., Grigsby, P. W., Cooper, J., Rotman, M., Gershenson, D., and Mutch, D. G. **Pelvic Irradiation With Concurrent Chemotherapy Versus Pelvic and Para-Aortic Irradiation for High-Risk Cervical Cancer: An Update of Radiation Therapy Oncology Group Trial (RTOG) 90-01**. *J Clin Oncol* 22 (2004), pp. 872–880.

Eminowicz, G., Hall-Craggs, M., Diez, P., and McCormack, M. **Improving target volume delineation in intact cervical carcinoma: Literature review and step-by-step pictorial atlas to aid contouring**. *Practical Radiation Oncology* 6.5 (2016), e203–e213.

Eminowicz, G. and Mccormack, M. **Cervical cancer radiotherapy Variability of clinical target volume delineation for definitive radiotherapy in cervix cancer**. *Radiotherapy and Oncology* 117 (2015), pp. 542–547.

Eminowicz, G., Rompokos, V., Stacey, C., and Mccormack, M. **The dosimetric impact of target volume delineation variation for cervical cancer radiotherapy**. *Radiotherapy and Oncology* 120 (2016), pp. 493–499.

Erickson, S. R., Chang, A., Johnson, C. E., and Gruppen, L. D. **Lecture versus Web tutorial for pharmacy students' learning of MDI technique.** *The Annals of pharmacotherapy* 37.4 (2003), pp. 500–5.

Esthappan, J., Ma, D. J., Narra, V. R., Raptis, C. A., and Grigsby, P. W. **Comparison of apparent diffusion coefficient maps to T2-weighted images for target delineation in cervix cancer brachytherapy**. *Journal of Contemporary Brachytherapy* 3.4 (2011), pp. 193–198.

Fairchild, A., Collette, L., Hurkmans, C. W., Baumert, B., Weber, D. C., Gulyban, A., and Poortmans, P. **Do results of the EORTC dummy run predict quality of radiotherapy delivered within multicentre clinical trials?** *European Journal of Cancer* 48 (2012), pp. 3232–3239.

Fairchild, A., Aird, E., Fenton, P. A., Gregoire, V., Gulyban, A., Lacombe, D., Matzinger, O., Poortmans, P., Ruyskart, P., Weber, D. C., and Hurkmans, C. W. **EORTC Radiation Oncology Group quality assurance platform: Establishment of a digital central review facility**. *Radiotherapy and Oncology* 103.3 (2012), pp. 279–286.

Fairchild, A., Langendijk, J. A., Nuyts, S., Scrase, C., Tomsej, M., Schuring, D., Gulyban, A., Ghosh, S., Weber, D. C., and Budach, W. **Quality assurance for the EORTC 22071-26071 study: dummy run prospective analysis** (2014).

Fairchild, A., Straube, W., Laurie, F., and Followill, D. **Does quality of radiation therapy predict outcomes of multicenter cooperative group trials? A literature review**. *International Journal of Radiation Oncology Biology Physics* 87.2 (2013), pp. 246–260.

Fairchild, A., Weber, D. C., Bar-Deroma, R., Gulyban, A., Fenton, P. A., Stupp, R., and Baumert, B. G. **Quality assurance in the EORTC 22033-26033/CE5 phase III randomized trial for low grade glioma: The digital individual case review**. *Radiotherapy and Oncology* 103 (2012), pp. 287–292.

Fenton, P. A., Hurkmans, C., Gulyban, A., Van Der Leer, J., Matzinger, O., Poortmans, P., Collette, L., and Bolla, M. **Quality assurance in prostate RT Quality assurance of the EORTC 22043-30041 trial in post-operative radiotherapy in prostate cancer: Results of the Dummy Run procedure**. *Radiotherapy and Oncology* 107 (2013), pp. 346–351.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D. M., Forman, D., and Bray, F. **Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012**. *International Journal of Cancer* 136.5 (2015), E359–E386. arXiv: `arXiv:1011.1669v3`.

FitzGerald, T. J. **What we have learned: the impact of quality from a clinical trials perspective**. *Seminars in Radiation Oncology* 22.1 (2012), pp. 18–28.

Fitzgerald, T. J. **A New Model for Imaging and Radiation Therapy Quality Assurance in the National Clinical Trials Network of the National Cancer Institute**. *Int J Radiat Oncol Biol Phys* 88.2 (2013), pp. 272–273.

Fokas, E., Clifford, C., Spezi, E., Joseph, G., Branagan, J., Hurt, C., Nixon, L., Abrams, R., Staffurth, J., and Mukherjee, S. **Comparison of investigator-delineated gross tumor volumes and quality assurance in pancreatic cancer: Analysis of the pretrial benchmark case for the SCALOP trial**. *Radiotherapy and Oncology* 117.3 (2015), pp. 432–437.

Foppiano, F., Fiorino, C., Frezza, G., Greco, C., and Valdagni, R. **The impact of contouring uncertainty on rectal 3D dose– volume data: results of a dummy run in a multicenter trial (AIROPROS01– 02)**. *2International Journal of Radiation Oncology Biology Physics* 57.2 (2003), pp. 573–579.

Fotina, I., Lütgendorf-Caucig, C., Stock, M., Pötter, R., and Georg, D. **Critical discussion of evaluation parameters for inter-observer variability in target definition for radiation therapy.** *Strahlentherapie und Onkologie* 188.2 (2012), pp. 160–167.

Gandhi, A. K., Sharma, D. N., Rath, G. K., Kumar Julka, P., Subramani, V., Sharma, S., Manigandan, D., Laviraj, M. A., Kumar, S., and Thulkar, S. **Clinical Investigation: Gynecologic Cancer Early Clinical Outcomes and Toxicity of Intensity Modulated Versus Conventional Pelvic Radiation Therapy for Locally Advanced Cervix Carcinoma: A Prospective Randomized Study**. *Int J Radiation Oncol Biol Phys* 87.3 (2013), pp. 542–548.

Gay, H. A., Barthold, H. J., O 'meara, E., Bosch, W. R., Naqa, I. E., Al-Lozi, R., Rosenthal, S. A., Lawton, C., Lee, W. R., Sandler, H., Zietman, A., Myerson, R., Dawson, L. A.,

Willett, C., Kachnic, L. A., Jhingran, A., Portelance, L., Ryu, J., Small, W., Gaffney, D., Viswanathan, A. N., and Michalski, J. M. **FEMALE PELVIS Normal Tissue RTOG Consensus Contouring Guidelines** (2011).

Gay, H. A., Barthold, H. J., O'Meara, E., Bosch, W. R., El Naqa, I., Al-Lozi, R., Rosenthal, S. A., Lawton, C., Lee, W. R., Sandler, H., Zietman, A., Myerson, R., Dawson, L. A., Willett, C., Kachnic, L. A., Jhingran, A., Portelance, L., Ryu, J., Small, W., Gaffney, D., Viswanathan, A. N., and Michalski, J. M. **Pelvic normal tissue contouring guidelines for radiation therapy: A radiation therapy oncology group consensus panel atlas**. *International Journal of Radiation Oncology Biology Physics* 83.3 (2012).

Georg, D., Kirisits, C., Hillbrand, M., Dimopoulos, J., and Tter, R. P. **Image-guided radiotherapy for cervix cancer: high-tech external beam therapy versus high-tech brachytherapy**. *Int. J. Radiation Oncology Biol. Phys.* 71.4 (2008), pp. 1272–1278.

Grau Eriksen, J., Salembier, C., Rivera, S., De Bari, B., Berger, D., Mantello, G., Müller, A.-C., Navarro Martin, A., Pasini, D., Tanderup, K., Palmu, M., Verfaillie, C., Pötter, R., and Valentini, V. **ESTRO FALCON e-learning Four years with FALCON – An ESTRO educational project: Achievements and perspectives on behalf of ESTRO**. *Radiotherapy and Oncology* 112 (2014), pp. 145–149.

Haie-Meder, C., Morice, P., and Castiglione, M. **Cervical cancer: ESMO clinical practice guidelines for diagnosis, treatment and follow-up**. *Annals of Oncology* 21.SUPPL. 5 (2010), pp. 2008–2011.

Haie-Meder, C., Morice, P., Castiglione, M., and Group, E. G. W. **Cervical cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up.** *Annals of oncology : official journal of the European Society for Medical Oncology* 21 Suppl 5 (2010), pp. v37–v40.

Haie-Meder, C., Pötter, R., Van Limbergen, E., Briot, E., De Brabandere, M., Dimopoulos, J., Dumas, I., Hellebust, T. P., Kirisits, C., Lang, S., Muschitz, S., Nevinson, J., Nulens, A., Petrow, P., and Wachter-Gerstner, N. **Recommendations from Gynaecological (GYN) GEC-ESTRO Working Group (I): Concepts and terms in 3D image based 3D treatment planning in cervix cancer brachytherapy with emphasis on MRI assessment of GTV and CTV**. *Radiotherapy and Oncology* 74.3 (2005), pp. 235–245.

Han, K., Milosevic, M., Fyles, A., Pintilie, M., and Viswanathan, A. N. **Clinical Investigation: Gynecologic Cancer Trends in the Utilization of Brachytherapy in**

Cervical Cancer in the United States. *Radiation Oncology Biology* 87.1 (2013), pp. 111–119.

Haworth, A., Kearvell, R., Greer, P. B., Hooton, B., Denham, J. W., Lamb, D., Duchesne, G., Murray, J., and Joseph, D. **Assuring high quality treatment delivery in clinical trials – Results from the Trans-Tasman Radiation Oncology Group (TROG) study 03.04 "RADAR " set-up accuracy study**. *Radiotherapy and Oncology* 90 (2009), pp. 299–306.

Hillemanns, P., Soergel, P., Hertel, H., and Jentschke, M. **Epidemiology and Early Detection of Cervical Cancer**. *Oncol Res Treat* 39 (2016), pp. 501–506.

Hoang Duc, A. K., Eminowicz, G., Mendes, R., Wong, S.-L., McClelland, J., Modat, M., Cardoso, M. J., Mendelson, A. F., Veiga, C., Kadir, T., D'Souza, D., and Ourselin, S. **Validation of clinical acceptability of an atlas-based segmentation algorithm for the delineation of organs at risk in head and neck cancer.** *Medical physics* 42.9 (2015), pp. 5027–34.

Hyeon Joo, J., Kim, Y. S., Cho, B. C., Jeong, C. Y., Park, W., Kim, H. J., Yoon, W. S., Yoon, M. S., Kim, J.-Y., Choi, J. H., Choi, Y., and Kim, J.-Y. **Variability in target delineation of cervical carcinoma: A Korean radiation oncology group study (KROG 15-06)**. *PLOS ONE* 12.3 (2017), pp. 1–12.

HYPOG-01, https://www.gustaveroussy.fr/en/node/3646.

Institute of Radiation Hygiene and World Heath Organization. **Quality Assurance in Radiotherapy. A guide prepared following a workshop held at Schloss Reisensburg, Federal Republic of Germany, 3-7 December 1984**. Tech. rep. Geneva: Institute of Radiation Hygiene - Federal Health Office (Neuherberg, Federal Republic of Germany) Organization, World Health Organization, 1988, p. 63.

Joura, E. A., Giuliano, A. R., Iversen, O.-E., Bouchard, C., Mao, C., Mehlsen, J., Moreira, E. D., Ngan, Y., Petersen, L. K., Lazcano-Ponce, E., Pitisuttithum, P., Restrepo, J. A., Stuart, G., Woelber, L., Yang, Y. C., Cuzick, J., Garland, S. M., Huh, W., Kjaer, S. K., Bautista, O. M., Chan, I. S. F., Chen, J., Gesser, R., Moeller, E., Ritter, M., Vuocolo, S., and Luxembourg, A. **A 9-valent HPV vaccine against infection and intraepithelial neoplasia in women.** *The New England journal of medicine* 372.8 (2015), pp. 711–723.

Kirisits, C., Federico, M., Nkiwane, K., Fidarova, E., Jürgenliemk-Schulz, I., De Leeuw, A., Lindegaard, J., Pötter, R., and Tanderup, K. **Cervical cancer brachytherapy Quality assurance in MR image guided adaptive brachytherapy for cervical cancer:**

Final results of the EMBRACE study dummy run. *Radiotherapy and Oncology* 117 (2015), pp. 548–554.

Kitchener, H. C., Fletcher, I., Roberts, C., Wheeler, P., Almonte, M., and Maguire, P. **The psychosocial impact of human papillomavirus testing in primary cervical screening - A study within a randomized trial**. *International Journal of Gynecological Cancer* 18.4 (2008), pp. 743–748.

Koh, W.-J., Greer, B. E., Abu-Rustum, N. R., Campos, S. M., Cho, K. R., Chon, H. S., Chu, C., Cohn, D., Crispens, M. A., Dizon, D. S., Dorigo, O., McMillian, N., Scavone, J., Eifel, P., Fisher, C. M., Frederick, P., Gaffney, D. K., Han, E., Higgins, S., Huh, W. K., Lurain, J. R., Mariani, A., Mutch, D., Nagel, C., Nekhlyudov, L., Nickles Fader, A., Remmenga, S. W., Reynolds, R. K., Tillmanns, T., Ueda, S., Valea, F. A., Emily, W., and Yashar, C. M. **NCCN Guidelines Version 1.2017 Panel Members Cervical Cancer NCCN Evidence Blocks**. Tech. rep. National Comprehensive Cancer Network, 2017.

Kouloulias, V. E. **Quality assurance in radiotherapy**. *European Journal of Cancer* 39 (2003), pp. 415–422.

Kouloulias, V. E., Giraud, J.-Y., Davis, B. J., Dusserre, A., Zurlo, A., and Bolla, M. **Quality assurance of the 22961 EORTC trial. A phase III study of the optimal combination of hormonal adjuvant treatment by LHRH analogue and radiation therapy for the management of locally advanced prostate cancer: the dummy run**. *Radiotherapy and Oncology* 73 (2004), pp. 11–20.

Kronz, J. D., Silberman, M. A., Allsbrook, W. C., and Epstein, J. I. **A web-based tutorial improves practicing pathologists' Gleason grading of images of prostate carcinoma specimens obtained by needle biopsy: validation of a new medical education paradigm.** *Cancer* 89.8 (2000), pp. 1818–23.

Kuznetsova, A., Bruun Brockhoff, P., and Haubo Bojesen Christensen, R. *lmerTest: Tests in Linear Mixed Effects Models*. 2016.

Lea, J. S. and Lin, K. Y. **Cervical Cancer**. *OGC* 39 (2012), pp. 233–253.

Lenth, R. V. **Least-Squares Means: The R Package lsmeans**. *Journal of Statistical Software* 69.1 (2016), pp. 1–33.

Lo, A. C., Liu, M., Chan, E., Lund, C., Truong, P. T., Loewen, S., Cao, J., Schellenberg, D., Carolan, H., Berrang, T., Wu, J., Berthelet, E., and Olson, R. **The impact of peer review of volume delineation in stereotactic body radiation therapy planning for primary lung cancer: a multicenter quality assurance study.** *Journal of tho-*

racic oncology : official publication of the International Association for the Study of Lung Cancer 9.4 (2014), pp. 527–33.

Lowenstein, J., Roll, J., Hanson, W., Davis, D., Lanciano, R., Calkins, A., Petereit, D., Varia, M., and Ibbott, G. **Radiotherapy Quality Assurance of Gynecologic Oncology Group (GOG) Protocol 165, A Cooperative Group Study of Carcinoma of the Cervix**. *International Journal of Radiation Oncology Biology Physics* 54.S1 (2002), p. 283.

Marnitz, S., Lukarski, D., Hler, C. K., Wlodarczyk, W., Ebert, A., Budach, V., Schneider, A., and Stromberger, C. **Helical tomotherapy versus conventional intensity-modulated radiation therapy forprimary chemoradiation in cervical cancer patients: an intraindividual comparison**. *Radiation Oncology Biology* 81 (2011), pp. 424–430.

Mayo, C., Moran, J., Xiao, Y., Bosch, W., Matuszak, M., Marks, L., Miller, R., Wu, Q., Yock, T., Popple, R., McNutt, T., Brown, N., Molineu, A., Purdie, T., Yorke, E., Santanam, L., Gabriel, P., Michalski, J., Moore, J., Richardson, S., Siochi, R., Napalitano, M., Feng, M., Fitzgerald, T., Ulin, K., Verbakel, W., Siddiqui, M., Martel, M., Archambault, Y., Morgas, T., Purcy, J., Adams, J., Ladra, M., Lansing, B., Ruo, R., Fogliata-Cozzi, A., and Hurkmans, C. **SU-E-P-22: AAPM Task Group 263 Tackling standardization of nomenclature for radiation therapy.** *International Journal of Radiation Oncology Biology Physics* 93.3S (2015), E383.

Mazeron, R., Gouy, S., Escande, A., Dumas, I., Rivin, E., Campo, d., Bentivegna, E., Bacorro, W., Lefkopoulos, D., Deutsch, E., Morice, P., Haie-Meder, C., and Chargari, C. **Locally advanced cervical cancer: Is it relevant to report image-guided adaptive brachytherapy using point A dose?** *Brachytherapy* 16.4 (2017), pp. 862–869.

Mcenery, K. W., Roth, S. M., Kelley, L. K., Hirsch, K. R., Menton, D. N., Kelly, E. A., and Washington, M. L. S. **A method for interactive medical instruction utilizing the World Wide Web**. *Proceedings. Symposium on Computer Applications in Medical Care* (1995), pp. 502–507.

McNemar, Q. **Note on the sampling error of the difference between correlated proportions or percentages**. *Psychometrika* 12.2 (1947), pp. 153–157.

Melidis, C., Bosch, W. R., Izewska, J., Fidarova, E., Zubizarreta, E., Ulin, K., Satoshi Ishikura, J., Followill, D., Physicist, M., Galvin, J., Haworth, A., Deidre Besuijen, Y., Clark Clark, z. H., Elizabeth Miles, X., Edwin Aird, J., Damien Weber, j. C., Hurkmans, C. W., Verellen, D., Kron, T., Ebert, M., Hatton, J., Gulyban, A., Corning, C., Giano-

lini, S., States, U., Michalski, J., Ibbott, G., Bhadrasain, V., Molineau, A., Straube, W., and Xiao, Y. **Global Harmonization of Quality Assurance Naming Conventions in Radiation Therapy Clinical Trials**. *International Journal of Radiation Oncology Biology Physics* 90.5 (2014), pp. 1242–1249.

Mendez, L. C., Leung, E., Cheung, P., and Barbera, L. **The Role of Stereotactic Ablative Body Radiotherapy in Gynaecological Cancers: A Systematic Review**. *Clinical Oncology* 29 (2017), pp. 378–384.

Miles, E. and Venables, K. **Radiotherapy Quality Assurance: Facilitation of Radiotherapy Research and Implementation of Technology**. *Clinical Oncology* 24.10 (2012), pp. 710–712.

Nakagawa, S. and Schielzeth, H. **Repeatability for Gaussian and non-Gaussian data: a practical guide for biologists.** *Biological reviews of the Cambridge Philosophical Society* 85.4 (2010), pp. 935–956.

Neoadjuvant Chemotherapy for Cervical Cancer Meta-analysis Collaboration. **Neoadjuvant chemotherapy for locally advanced cervical cancer: a systematic review and meta-analysis of individual patient data from 21 randomised trials**. *European Journal of Cancer* 39 (2003), pp. 2470–2486.

Nilsson, P., Nilsson, P., Ceberg, C., Kjell É N, E., Gagliardi, G., Blomgren, K., Nilsson, S., Johansson, M., and Glimelius, B. **A template for writing radiotherapy protocols**. *Acta Oncologica* 54.2 (2014), pp. 275–279.

Ohri, N., Shen, X., Dicker, A. P., Doyle, L. A., Harrison, A. S., and Showalter, T. N. **Radiotherapy protocol deviations and clinical outcomes: A meta-analysis of cooperative group clinical trials**. *Journal of the National Cancer Institute* 105.6 (2013), pp. 387–393.

Pearcey, R., Miao, Q., Kong, W., Zhang-Salomons, J., and Mackillop, W. J. **Impact of Adoption of Chemoradiotherapy on the Outcome of Cervical Cancer in Ontario: Results of a Population- Based Cohort Study**. *J Clin Oncol* 25 (2007), pp. 2383–2388.

Peters, L. J., O'Sullivan, B., Giralt, J., Fitzgerald, T. J., Trotti, A., Bernier, J., Bourhis, J., Yuen, K., Fisher, R., and Rischin, D. **Critical impact of radiotherapy protocol compliance and quality in the treatment of advanced head and neck cancer: Results from TROG 02.02**. *Journal of Clinical Oncology* 28.18 (2010), pp. 2996–3001.

Peters, W. A., Liu, P. Y., Barrett, R. J., Stock, R. J., Monk, B. J., Berek, J. S., Souhami, L., Grigsby, P., Gordon, W., and Alberts, D. S. **Concurrent chemotherapy and pelvic**

radiation therapy compared with pelvic radiation therapy alone as adjuvant therapy after radical surgery in high-risk early-stage cancer of the cervix. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 18.8 (2000), pp. 1606–1613.

Petersen, R. P., Truong, P. T., Kader, H. A., Berthelet, E., Lee, J. C., Hilts, M. L., Kader, A. S., Beckham, W. A., and Olivotto, I. A. **Target Volume Delineation for Partial Breast Radiotherapy Planning: Clinical Characteristics Associated with Low Interobserver Concordance**. *International Journal of Radiation Oncology Biology Physics* 69.1 (2007), pp. 41–48.

Petric, P., Dimopoulos, J., Kirisits, C., Berger, D., Hudej, R., and Pötter, R. **Inter- and intraobserver variation in HR-CTV contouring: Intercomparison of transverse and paratransverse image orientation in 3D-MRI assisted cervix cancer brachytherapy**. *Radiotherapy and Oncology* 89.2 (2008), pp. 164–171.

Petrič, P., Hudej, R., Rogelj, P., Blas, M., Tanderup, K., Fidarova, E., Kirisits, C., Berger, D., Carl, J., Dimopoulos, A., Pötter, R., and Hellebust, T. P. **Uncertainties of target volume delineation in MRI guided adaptive brachytherapy of cervix cancer: A multi-institutional study**. *Radiotherapy and Oncology* 107 (2013), pp. 6–12.

Pike, M. C., Pearce, C. L., and Wu, A. H. **Prevention of cancers of the breast, endometrium and ovary**. *Oncogene* 23, (2004), pp. 6379–6391.

Plummer, M., Peto, J., and Franceschi, S. **Time since first sexual intercourse and the risk of cervical cancer**. *International Journal of Cancer* 130.11 (2012), pp. 2638–2644. arXiv: `NIHMS150003`.

Poortmans, P. M. P., Davis, J. B., Ataman, F., Bernier, J., and Horiot, J.-C. **The quality assurance programme of the Radiotherapy Group of the European Organisation for Research and Treatment of Cancer: past, present and future for the EORTC Radiotherapy Group**. *Eur J Surg Oncol* 31.6 (2005), pp. 667–74.

Poortmans, P. M. P., Collette, S., Kirkove, C., Van Limbergen, E., Budach, V., Struikmans, H., Collette, L., Fourquet, A., Maingon, P., Valli, M., De Winter, K., Marnitz, S., Barillot, I., Scandolaro, L., Vonk, E., Rodenhuis, C., Marsiglia, H., Weidner, N., van Tienhoven, G., Glanzmann, C., Kuten, A., Arriagada, R., Bartelink, H., and Van den Bogaert, W. **Internal Mammary and Medial Supraclavicular Irradiation in Breast Cancer**. *New England Journal of Medicine* 373 (2015), pp. 317–27.

Poortmans, P. M. P., Kouloulias, V., Van Tienhoven, G., Collette, L., Struikmans, H., Venselaar, J. L. M., Van Den Bogaert, W., Davis, J. B., and Lambin, P. **Strahlentherapie und**

Onkologie Quality Assurance in the EORTC Randomized Trial 22922/10925 Investigating the Role of Irradiation of the Internal Mammary and Medial Supraclavicular Lymph Node Chain Works , on Behalf of the EORTC Radiation Oncology and Brea. *Strahlenther Onkol* 182.10 (2006), pp. 576–82.

Poortmans, P. M. P., Venselaar, J. L. M., Struikmans, H., Hurkmans, C. W., Davis, J. B., Huyskens, D., Tienhoven Van, G., Vlaun, V., Lagendijk, J. J. W., Mijnheer, B. J., De Winter, K. A., Van Der Hulst, M. H., and Van Den Bogaert, W. F. C. **The potential impact of treatment variations on the results of radiotherapy of the internal mammary lymph node chain: a quality-assurance report on the dummy run of EORTC phase III randomized trial 22922/10925 in stage I–III breast cancer**. *International Journal of Radiation Oncology Biology Physics* 49.5 (2001), pp. 1399–1408.

Pötter, R., Georg, P., Dimopoulos, J. C. A., Grimm, M., Berger, D., Nesvacil, N., Georg, D., Schmid, M. P., Reinthaller, A., Sturdza, A., and Kirisits, C. **Clinical outcome of protocol based image (MRI) guided adaptive brachytherapy combined with 3D conformal radiotherapy with or without chemotherapy in patients with locally advanced cervical cancer**. *Radiotherapy and Oncology* 100.1 (2011), pp. 116–123.

Pötter, R., Haie-Meder, C., Van Limbergen, E., Barillot, I., De Brabandere, M., Dimopoulos, J., Dumas, I., Erickson, B., Lang, S., Nulens, A., Petrow, P., Rownd, J., and Kirisits, C. **Recommendations from gynaecological (GYN) GEC ESTRO working group (II): Concepts and terms in 3D image-based treatment planning in cervix cancer brachytherapy - 3D dose volume parameters and aspects of 3D image-based anatomy, radiation physics, radiobiolo**. *Radiotherapy and Oncology* 78.1 (2006), pp. 67–77.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria, 2016.

Ramlov, A., Assenholt, M. S., Jensen, M. F., Grønborg, C., Nout, R., Alber, M., Fokdal, L., Tanderup, K., and Lindegaard, J. C. **Clinical implementation of coverage probability planning for nodal boosting in locally advanced cervical cancer**. *Radiotherapy and Oncology* 123 (2017), pp. 158–163.

Rasch, C., Barillot, I., Remeijer, P., Touw, A., Van Herk, M., and Lebesque, J. V. **Definition of the prostate in CT and MRI: A multi-observer study**. *International Journal of Radiation Oncology Biology Physics* 43.1 (1999), pp. 57–66.

Reuzé, S., Orlhac, F., Chargari, C., Nioche, C., Limkin, E., Riet, F., Escande, A., Haie-meder, C., Gouy, S., Buvat, I., Deutsch, E., and Robert, C. **Prediction of cervical cancer recurrence using textural features extracted from 18 F-FDG PET images acquired with different scanners**. *Oncotarget* 8.26 (2017), pp. 43169–43179.

Ridgway, P. F., Sheikh, A., Sweeney, K. J., Evoy, D., McDermott, E., Felle, P., Hill, A. D., and O'Higgins, N. J. **Surgical e-learning: Validation of multimedia web-based lectures**. *Medical Education* 41.2 (2007), pp. 168–172.

Rivin del Campo, E., Rivera, S., Martínez-Paredes, M., Hupé, P., Slocker Escarpa, A., Borget, I., Mazeron, R., Scholl, S., Palacios Eito, A., Haie-Meder, C., Chargari, C., and Deutsch, E. **Assessment of the novel online delineation workshop dummy run approach using FALCON within a European multicentre trial in cervical cancer (RAIDs).** *Radiotherapy and oncology : journal of the European Society for Therapeutic Radiology and Oncology* 124.1 (2017), pp. 130–138.

Ronco, G., Dillner, J., Miriam Elfström, K., Tunesi, S., F Snijders, P. J., Arbyn, M., Kitchener, H., Segnan, N., Gilham, C., Giorgi-Rossi, P., Berkhof, J., Peto, J., and L M Meijer, C. J. **Efficacy of HPV-based screening for prevention of invasive cervical cancer: follow-up of four European randomised controlled trials**. *The Lancet* 383 (2014), pp. 524–532.

Rose, P. G., Ali, S., Watkins, E., Thigpen, J. T., Deppe, G., Clarke-Pearson, D. L., and Insalaco, S. **Long-Term Follow-Up of a Randomized Trial Comparing Concurrent Single Agent Cisplatin, Cisplatin-Based Combination Chemotherapy, or Hydroxyurea During Pelvic Irradiation for Locally Advanced Cervical Cancer: A Gynecologic Oncology Group Study**. *J Clin Oncol* 25 (2007), pp. 2804–2810.

Samson, K. K., Haynatzki, G., Soliman, A. S., and Valerianova, Z. **Temporal changes in the cervical cancer burden in Bulgaria: Implications for eastern european countries going through transition**. *Cancer Epidemiol* 44 (2016), pp. 154–160.

Sant, M., Dolores, M., Lopez, C., Agresti, R., Sánchez Pérez, M. J., Holleczek, B., Bielska-Lasota, M., Dimitrova, N., Innos, K., Katalinic, A., Langseth, H., Larrañ Aga, N., Rossi, S., Siesling, S., Minicozzi, P., The, E.-5., and Group, W. **Survival of women with cancers of breast and genital organs in Europe 1999–2007: Results of the EUROCARE-5 study**. *Z. Valerianova, N. Dimitrova (Bulgarian CR)Czech National CRDanish Cancer Society Tallinn) N. Malila, K. Seppä (Finnish CR* Finland (2015).

Sanuki-Fujimoto, N., Ishikura, S., Hayakawa, K., Kubota, K., Nishiwaki, Y., and Tamura, T. **Radiotherapy quality assurance review in a multi-center randomized trial**

of limited-disease small cell lung cancer: the Japan Clinical Oncology Group (JCOG) trial 0202. *Radiation Oncology* 4.16 (2009).

Schefter, T., Winter, K., Kwon, J. S., Stuhr, K., Balaraj, K., Yaremko, B. P., Small Jr, W., and Gaffney, D. K. **Clinical Investigation: Gynecologic Cancer A Phase II Study of Bevacizumab in Combination With Definitive Radiotherapy and Cisplatin Chemotherapy in Untreated Patients With Locally Advanced Cervical Carcinoma: Preliminary Results of RTOG 0417**. *Radiation Oncology Biology* 83.4 (2012), pp. 1179–1184.

Schefter, T., Winter, K., Kwon, J. S., Stuhr, K., Balaraj, K., Yaremko, B. P., Small, W., Sause, W., and Gaffney, D. **Clinical Investigation: Gynecologic Cancer RTOG 0417: Efficacy of Bevacizumab in Combination With Definitive Radiation Therapy and Cisplatin Chemotherapy in Untreated Patients With Locally Advanced Cervical Carcinoma**. *Radiation Oncology Biology* 88.1 (2014), pp. 101–105.

Sharpe, R., Benfield, G., Roberts, G., and Francis, R. **The undergraduate experience of blended e-learning: a review of UK literature and practice Executive summary** (2006), pp. 1–103.

Spoelstra, F. O. B., Senan, S., Le, C., Choux, P., Ishikura, S., Casas, F., Ball, D., Price, A., De Ruysscher, D., Van, J. R., Rnsen, S., and Koste, D. **Variations in target volume definition for postoperative radiotherapy in stage III non–small-cell lung cancer: analysis of an international contouring study**. *International Journal of Radiation Oncology Biology Physics* 76.4 (2010), pp. 1106–1113.

Spry, N., Bydder, S., Harvey, J., Borg, M., Ngan, S., Millar, J., Graham, P., Zissiadis, Y., Kneebone, A., and Ebert, M. **Accrediting radiation technique in a multicentre trial of chemoradiation for pancreatic cancer**. *Journal of Medical Imaging and Radiation Oncology* 52.6 (2008), pp. 598–604.

Stehman, F. B., Ali, S., Keys, H. M., Muderspach, L. I., Chafe, W. E., Gallup, D. G., Walker, J. L., Gersell, D., Author, C., Clarence Ehrlich Professor, T. E., and Mackey, D. **Radiation therapy with or without weekly cisplatin for bulky stage ib cervical carcinoma; follow-up of a gynecologic oncology group trial**. *Am J Obstet Gynecol* 197.5 (2007), pp. 503–1.

Sturdza, A., Pötter, R., Ulrik Fokdal, L., Haie-Meder, C., Tee Tan, L., Mazeron, R., Petric, P., Šegedin, B., Maria Jurgenliemk-Schulz, I., Nomden, C., Gillham, C., McArdle, O., Van Limbergen, E., Janssen, H., Hoskin, P., Lowe, G., Tharavichitkul, E., Villafranca, E.,

Mahantshetty, U., Georg, P., Kirchheiner, K., Kirisits, C., Tanderup, K., and Christian Lindegaard, J. **Image guided brachytherapy in cervical cancer Image guided brachytherapy in locally advanced cervical cancer: Improved pelvic control and survival in RetroEMBRACE, a multicenter cohort study**. *Radiotherapy and Oncology* 120 (2016), pp. 428–433.

Taha, A. A. and Hanbury, A. **Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool**. *BMC Medical Imaging* 15 (2015), p. 29.

Tanderup, K., Fokdal, L. U., Sturdza, A., Haie-Meder, C., Mazeron, R., van Limbergen, E., Jürgenliemk-Schulz, I., Petric, P., Hoskin, P., Dörr, W., Bentzen, S. M., Kirisits, C., Lindegaard, J. C., and Pötter, R. **Effect of tumor dose, volume and overall treatment time on local control after radiochemotherapy including MRI guided brachytherapy of locally advanced cervical cancer**. *Radiotherapy and Oncology* (2016).

Taylor, A., Rockall, A. G., and Powell, M. E. B. **An Atlas of the Pelvic Lymph Node Regions to Aid Radiotherapy Target Volume Definition**. *Clinical Oncology* 19.7 (2007), pp. 542–550.

Taylor, A., Rockall, A. G., Reznek, R. H., and Powell, M. E. B. **Mapping pelvic lymph nodes: Guidelines for delineation in intensity-modulated radiotherapy**. *International Journal of Radiation Oncology Biology Physics* 63.5 (2005), pp. 1604–1612.

Thwaitesa, D., Scalliefb, P., Leer, J. W., and Overgaardd, J. **Quality assurance in radiotherapy**. *Radiotherapy and Oncology* 35 (1995), pp. 61–73.

Toita, T., Oguchi, M., Ohno, T., Kato, S., Niibe, Y., Kodaira, T., Kazumoto, T., Kataoka, M., Shikama, N., Kenjo, M., Teshima, T., and Kagami, Y. **Quality Assurance in the Prospective Multi-institutional Trial on Definitive Radiotherapy Using High-dose-rate Intracavitary Brachytherapy for Uterine Cervical Cancer: The Individual Case Review**. *Jpn J Clin Oncol* 39.12 (2009), pp. 813–819.

Vale, C. **Reducing Uncertainties About the Effects of Chemoradiotherapy for Cervical Cancer: A Systematic Review and Meta-Analysis of Individual Patient Data From 18 Randomized Trials**. *J Clin Oncol* 26 (2008), pp. 5802–5812.

Van der Schueren, E., Horiot, J. C., Leunens, G., Rubens, R., Steward, W., van Dongen, J. A., van Oosterom, A. T., and Vantongelen, K. **Quality assurance in cancer treatment. Report of a Working Party from the European School of Oncology.** *European journal of cancer (Oxford, England : 1990)* 29A.2 (1993), pp. 172–81.

Van Mourik, A. M., Elkhuizen, P. H. M., Minkema, D., Duppen, J. C., and van Vliet-Vroegindeweij, C. **Multiinstitutional study on target volume delineation variation in breast radiotherapy in the presence of guidelines**. *Radiotherapy and Oncology* 94.3 (2010), pp. 286–291.

Vinod, S. K., Jameson, M. G., Min, M., and Holloway, L. C. **Uncertainties in volume delineation in radiation oncology: A systematic review and recommendations for future studies**. *Radiotherapy and Oncology* 121.2 (2016), pp. 169–179.

Vinod, S. K., Lim, K., Bell, L., Veera, J., Ohanessian, L., Juresic, E., Borok, N., Chan, P., Chee, R., Do, V., Govindarajulu, G., Sridharan, S., Johnson, C., Moses, D., Van Dyk, S., and Holloway, L. **High-risk CTV delineation for cervix brachytherapy: Application of GEC-ESTRO guidelines in Australia and New Zealand**. *Journal of Medical Imaging and Radiation Oncology* 61.1 (2017), pp. 133–140.

Vordermark, D. **Radiotherapy of Cervical Cancer External-Beam Radiotherapy, Brachytherapy, Cisplatin: Elements of Standard Treatment**. *Oncol Res Treat* 39 (2016), pp. 516–520.

Weber, D. C., Hurkmans, C. W., Melidis, C., Budach, W., Langendijk, J. H., Peters, L. J., Grégoire, V., Maingon, P., Combescure, C., and ja Langendijk. **Outcome impact and cost-effectiveness of quality assurance for radiotherapy planned for the EORTC 22071-24071 prospective study for head and neck cancer**. *Radiotherapy and Oncology* 111 (2014), pp. 393–399.

Weber, D. C., Poortmans, P. M. P., Hurkmans, C. W., Aird, E., Gulyban, A., and Fairchild, A. **Quality assurance for prospective EORTC radiation oncology trials: The challenges of advanced technology in a multicenter international setting**. 2011.

Weber, D. C., Tomsej, M., Melidis, C., and Hurkmans, C. W. **QA makes a clinical trial stronger: Evidence-based medicine in radiation therapy**. *Radiotherapy and Oncology* 105.1 (2012), pp. 4–8.

Weiss, E., Richter, S., Krauss, T., Metzelthin, S. I., Hille, A., Pradier, O., Siekmeyer, B., Vorwerk, H., and Hess, C. F. **Conformal radiotherapy planning of cervix carcinoma: differences in the delineation of the clinical target volume. A comparison between gynaecologic and radiation oncologists**. *Radiotherapy and Oncology* 67 (2003), pp. 87–95.

White, J. E., Chen, T., Mccracken, J., Kennedy, P., Seydel, H. G., Hartman, G., Mira, J., Khan, M., Durrance, F. Y., and Skinner, O. **The Influence of Radiation Therapy Quality Control on Carcinoma of the Lung**. *Cancer* 50 (1982), pp. 1084–1090.

Whitney, C. W., Sause, W., Bundy, B. N., Malfetano, J. H., Hannigan, E. V., Fowler, W. C., Clarke-Pearson, D. L., and Liao, S.-Y. **Randomized Comparison of Fluorouracil Plus Cisplatin Versus Hydroxyurea as an Adjunct to Radiation Therapy in Stage IIB-IVA Carcinoma of the Cervix With Negative Para-Aortic Lymph Nodes: A Gynecologic Oncology Group and Southwest Oncology Group Study**. *J Clin Oncol* 17.5 (1999), pp. 1339–1348.

Wickham, H. **ggplot2: Elegant Graphics for Data Analysis**. Springer-Verlag New York, 2009.

Willett, C. G., Moughan, J., Galvin, J. M., Sc, D., Crane, C. H., Winter, K., Manfredi, D., Rich, T. A., Rabinovitch, R., Lustig, R., Machtay, M., and Curran Jr, W. J. **Compliance with Therapeutic Guidelines in Radiation Therapy Oncology Group Prospective Gastrointestinal Clinical Trials**. *Radiotherapy and Oncology* 105.1 (2012), pp. 9–13.

Wu, D. H., Mayr, N. A., Karatas, Y., Karatas, R., Adli, M., Edwards, S. M., Wolff, J. D., Movahed, A., Montebello, J. F., and Yuh, W. T. C. **Interobserver variation in cervical cancer tumor delineation for image-based radiotherapy planning among and within different specialties.** *Journal of applied clinical medical physics / American College of Medical Physics* 6.4 (2005), pp. 106–10.

Zou, K. H., Warfield, S. K., Bharatha, A., Tempany, C. M. C., Kaus, M. R., Haker, S. J., Wells, W. M., Jolesz, F. A., and Kikinis, R. **Statistical validation of image segmentation quality based on a spatial overlap index.** *Academic radiology* 11.2 (2004), pp. 178–189.