A review of uncertainties in radiotherapy dose reconstruction and their impacts on dose–response relationships

To cite this article: Jérémi V Bezin et al 2017 J. Radiol. Prot. 37 R1

View the article online for updates and enhancements.
A review of uncertainties in radiotherapy dose reconstruction and their impacts on dose–response relationships

Jérémy Vu Bezin, Rodrigue S Allodji, Jean-Pierre Mége, Guillaume Beldjoudi, Fleur Saunier, Jean Chavaudra, Eric Deutsch, Florent de Vathaire, Valérie Bernier, Christian Carrie, Dimitri Lefkopoulos and Ibrahima Diallo

1 Inserm, Radiation Epidemiology Team, CESP-U1018, F-94807, Villejuif, France
2 Gustave Roussy, Villejuif, F-94805, France
3 Paris-Sud University, Orsay, F-91400, France
4 Centre Léon Bérard, Lyon, F-69008, France
5 Institut de Cancérologie de Lorraine, Vandœuvre-lès-Nancy, F-54511, France
6 Inserm, UMR 1030, Villejuif, F-94807, France

E-mail: ibrahim.diallo@gustaveroussy.fr

Received 1 June 2016, revised 30 December 2016
Accepted for publication 6 January 2017
Published 24 January 2017

Abstract

Proper understanding of the risk of radiation-induced late effects for patients receiving external photon beam radiotherapy requires the determination of reliable dose–response relationships. Although significant efforts have been devoted to improving dose estimates for the study of late effects, the most often questioned explanatory variable is still the dose. In this work, based on a literature review, we provide an in-depth description of the radiotherapy dose reconstruction process for the study of late effects. In particular, we focus on the identification of the main sources of dose uncertainty involved in this process and summarise their impacts on the dose–response relationship for radiotherapy late effects. We provide a number of recommendations for making progress in estimating the uncertainties in current studies of radiotherapy late effects and reducing these uncertainties in future studies.

Keywords: uncertainties, dose estimates, dose–response relationship, radiotherapy late effects

(Some figures may appear in colour only in the online journal)
Figure 1. Schematic of the dose reconstruction process and exposure calculation. Dose reconstruction requires three types of input data: patient anatomy, treatment plan and characterisation of the irradiation source. Since each type of data information available in the patient treatment record may be missing, questions regarding completeness have been included in the schema. Two extreme cases have been given as answers to these questions. Uncertainty is then added to the dose reconstruction process according to the accuracy of the chosen answer. Finally, a three-dimensional dose matrix is obtained, which in turn is converted into an exposure value which will be used in the subsequent statistical study to characterise the dose of ionising radiation received by the patient.
1. Introduction

Proper understanding of the risk of radiation-induced late effects for patients undergoing external photon beam radiotherapy (RT) requires the determination of reliable dose–response relationships, which in turn—for patients with available long-term follow-up data—requires the retrospective estimation of the radiation dose distribution in the patient’s body [1, 2]. Many epidemiological studies on the health effects of radiation exposure derive risk estimates using confidence intervals that express only the impact of statistical fluctuations of the different data in the frame of the chosen risk model [3]. Other specific sources of uncertainty, such as those due to missing data needed for dose reconstructions (i.e. irradiation source characteristics, patient anatomy, treatment plan), those due to dose evaluation and those due to computation of the exposure value, representing the absorbed dose at a specific site, must also be addressed in order to clarify the current state of knowledge on dose–response relationships [4].

As illustrated in figure 1, risk estimation in studies of the late effects of RT can be described as a two-step process. The first step is the dose reconstruction [1, 5–7], which involves evaluation by medical physicists of the absorbed dose distribution inside the patient’s body. Since treatment often occurred many years earlier, data on imaging of patient anatomy, RT treatment plan and irradiation source may be missing and approximations will be necessary to proceed to dose evaluation. These approximations, along with inaccuracies in dose calculations and/or measurements, will translate into uncertainty in the dose evaluated in each volume element (voxel) forming the three-dimensional (3D) dose matrix. In the second step, in agreement with epidemiologists or biostatisticians, medical physicists use the dose matrix to derive an exposure index at the volume of interest [8, 9]. The exposure is defined according to the design of the late effects study. It is often a point dose or an average dose representing the exposure of a substructure or a whole organ. Thus, computation of the exposure value adds uncertainty to dose–response studies because a point dose, or an average dose, does not describe the dose distribution throughout the whole volume of interest.

If no efforts are made to account for uncertainties in dose assessment and exposure computation, they may affect the dose–response evaluation in several ways [10]. The estimated risk coefficients and their confidence limits can be biased, usually through underestimation [10]. Most epidemiological studies have limited power to make a distinction between linear and non-linear dose–response relationships, and the presence of such uncertainties adds to the difficulty [10].

In the present work, we reviewed the main sources of uncertainty involved in the process of retrospective reconstruction of organ doses in cohorts of patients treated by external photon beam RT. These were: (i) imaging of patient anatomy, (ii) reconstruction of the RT treatment plan, (iii) characterisation of the irradiation source, and (iv) measurement or calculation of the dose distributions. Then we summarised the impacts of these uncertainties on the published literature on the dose–response relationship in late effects of RT.

2. Sources of uncertainty in retrospective dose reconstruction for studies of the late effects of RT

2.1. Imaging of patient anatomy

Epidemiological studies require knowledge of patient anatomy from the treated volume to the volume where the outcome was observed. Nowadays, in RT, a patient’s anatomy is acquired by performing a computed tomography (CT) scan of the patient in the treatment position,
known as the RT-planning CT. The RT-planning CT is quite systematically limited to the region directly involved in the irradiation: the target volumes and organs at risk. If dose estimation is needed for more remote tissues or if the patient’s record does not include the RT-planning CT, anthropomorphic phantoms will be necessary. Anthropomorphic phantoms can be physical or virtual and represent an average person. In its simplest form, medical physicists have used a water tank with dimensions similar to those of an adult patient [11–17]. However, to limit the discrepancies between patient anatomy and the phantom, several phantoms with different anatomies will be necessary. Choices of physical phantoms are limited to a few anatomies with respect to the age (from newborn to adult) and gender (for adult phantoms only) of the patient [18–23]. They can usually simulate three densities of human body tissue: lungs, soft tissues and bones. Virtual phantoms can be further adapted to better match patient dimensions. As well as age and gender, height and weight can also be taken into account [1, 6, 24, 25]. Ideally, the size and position of the organs in virtual phantoms may be chosen with regard to anthropometric characteristics [6, 26–28]. Virtual phantoms can be used alone or in addition to a patient’s RT-planning CT [5].

For instance, to reconstruct the absorbed dose in active bone marrow for 15 patients, Veres et al [6] used whole-body RT-planning CT images from six patients as surrogate anatomies. The closest surrogate anatomy was matched for each patient by taking into account gender, age, weight, height and treatment position. When available, the external dimensions of the patient were used to scale the CT images to the patient’s anatomy. Subsequently, the distribution of active bone marrow in the skeleton was estimated using tabulations from the International Commission on Radiological Protection [29] according to patient age. Hence, doses could be calculated in a series of points distributed along the skeleton in such a way that their distribution reflected age-specific active bone marrow density.

However, only a few anthropomorphic phantoms have been adapted to different RT treatment positions [25]. Instead, medical physicists have used RT-planning CT of representative patients in specific treatment positions [7, 30–33]. In this case, organ contouring must be redone so as to include all organs of interest. Also, to personalise the phantom to the patient, it is possible to mix the representative patient image with virtual modelled organs [34]. The modelled organ is voxelised and inserted into the representative patient RT-planning CT, hence producing a hybrid phantom.

Uncertainties arising from simulation of the patient’s anatomy depend on the available data. If an RT-planning CT of the regions of interest is available, uncertainties predominantly arise from determination of the organ’s volume (inter- and intra-operator variabilities and partial volume effect) and its position during treatment. If an anthropomorphic phantom is used, the uncertainties are mainly due to discrepancies between the phantom anatomy and the patient’s true anatomy.

Fiorino et al [35] assessed inter-operator variability by asking five well-trained radiation therapists to contour the prostate and the seminal vesicles on CT images from six patients. The percentage standard deviation of the contoured volumes ranged from 10% to 18%. The intra-operator variability, which was estimated by asking the radiation therapists to contour the volumes of one patient for a second time, averaged 5%. Veres et al [26] estimated inter- and intra-operator variabilities for the thyroid using a Livermore phantom RS-550 (RSD Inc., Long Beach, CA, USA). The real thyroid volume was estimated by filling the Livermore thyroid with water. The phantom was CT-scanned and contoured by six different operators to estimate inter-operator variability, and contoured four times by a single operator to estimate intra-operator variability. The resulting variabilities were 6% and 3% for inter-operator and intra-operator variability, respectively. In the same work, Veres et al [26] also revealed a
systematic error on organ contouring which affected volume measurement. The authors concluded that this error was mainly due to the partial volume effect, but the contrast of visualisation, the calculation and display mode of the 3D structures used by the software could also be involved [36]. By modelling this systematic error, Veres et al found it was proportional to the CT slice thickness and inversely proportional to the volume of the organ. Hence, for a slice thickness of 2 mm, the error in thyroid volume was less than 5%.

Van Herk et al [37] quantified the motion of the prostate and seminal vesicles for 11 patients treated for prostate carcinoma. For each patient, four CT scans were made during the course of the treatment. For each CT scan, the volume of a region including the prostate and the seminal vesicles was contoured. The movements of the volume between the first scan and the following ones were quantified by the translation and rotation along three axes: cranial–caudal, left–right and anterior–posterior. The most important movements, when considering all patients, were anterior–posterior translations and rotations around the left–right axis with a magnitude of 2.7 mm and 4.0°, respectively. Other displacements were much smaller. Meijer et al [38] evaluated the organ motion of the bladder and setup errors during RT for bladder cancer. They reported that organ motion was the predominant geometric uncertainty in the RT process with an average shift of 5 mm along the cranial–caudal axis.

To evaluate the impact on dose evaluation of anatomical discrepancies between representative patient imaging and an actual patient, Taylor et al [30] planned four breast cancer treatments on 20 random patient RT-planning CTs. They calculated the mean dose to the heart in each case and the coefficient of variation for each treatment. The mean heart dose was between 1 Gy and 2 Gy with a coefficient of variation of 11% for right-sided tangential pair RT, between 2 Gy and 4 Gy (30%) for left-sided tangential pair RT, between 5 Gy and 15 Gy (21%) for right-sided internal mammary chain (IMC) RT and between 20 Gy and 29 Gy (11%) for left-sided IMC RT. The authors identified the patients’ anatomical variations as the greatest source of uncertainty when using a representative patient RT-planning CT to estimate the dose to the heart and coronary arteries during breast cancer RT. Besides patient anatomy, Taylor et al also evaluated the impact of patient position during treatment. A left tangential breast cancer treatment was reconstructed on the same patient for two treatment positions. First with both arms above the head and second with the ipsilateral arm abducted to 90° and the contralateral arm by the patient’s side. The calculated mean heart dose was 2 Gy and 3.4 Gy, respectively, hence a 70% increase, thereby indicating that, like patient anatomy, patient position can be an important parameter for dose reconstruction.

Scarboro et al [39] investigated the impact of the size and the position of an out-of-field organ, the stomach, on the mean dose it received during mantle field irradiation for Hodgkin lymphoma. To simulate the variability of the volume and position of the stomach among the population of an epidemiological study, the stomach volume was expanded from 433.77 cm³ to 2641.40 cm³ (600% volume increase; hence, at most, a 3 cm isotropic expansion) and it was successively shifted by about 3 cm in the medial, posterior, anterior, superior and inferior directions. These authors reported that shifts in stomach position toward or away from the treatment field significantly altered the mean stomach dose, between 1.43 Gy (2.5 cm inferior) and 2.76 Gy (2.5 cm superior), and that most variations in organ size had no significant impact on the mean dose, between 2.06 Gy (0 cm expansion) and 2.47 Gy (3 cm expansion). Among the nine tested stomach configurations, the coefficient of variation of the mean dose was about 15%. Hence, they concluded that these results could potentially impact the design of epidemiological studies of a radiation-induced late effects for organs that are known to vary in size and/or position between individuals. Their results also indicated that using the mean dose as an exposure value to characterise the stomach dose during Hodgkin lymphoma
mantle field irradiation could lead to a 70% relative uncertainty (3 cm expansion) due to local dose variability throughout the stomach volume, with an average around 50%.

To better adjust anatomical models to patient anatomy, virtual phantoms can be generated according to a patient’s anthropometric data. Veres et al [26] and Badouna et al [27] modelled the variability of the thyroid and heart volume and revealed that the body surface area was a suitable indicator for predicting organ volume. In comparison with phantom selection based on patient age, phantom selection based on body surface area could improve heart volume prediction by about 20% [27]. Lamart et al [28] developed multivariate linear regression models to predict the location and size of the stomach using predictor variables such as body mass index, ponderal index and age. To test their model, they generated a 3D computational stomach in a hybrid phantom for three selected patients and were able to derive good dose estimates with a difference of less than 10%–30% in comparison with dose estimates computed using images of the actual patient. Moignier et al [40] studied the coronary dose variations throughout 22 different coronary topologies and two breast RT beam setups. Twenty-two detailed heart models were created from heart CT-angiographies and inserted into representative patients’ RT-planning CTs. The highest variations were found for irradiation of tangential, tumoural bed and IMC beam setups. They reported that the coronary mean dose variation could be up to 40% between the different coronary topologies, whereas the mean heart dose variation was only up to 6%. These results suggest that large dose uncertainty exists if anatomical information is scarce, but also that selection of the volume of interest can be crucial for epidemiological study design.

Finally, tissue heterogeneity can be of interest in certain situations. Lamart et al [41] reported that failure to take into account tissue heterogeneities could lead to underestimation of the dose to the oesophagus after RT for breast cancer. In the case of a 220 kV orthovoltage irradiation, the dose to the cervical oesophagus could be underestimated by approximately 15% because of the presence of the trachea anterior to the oesophagus. In the case of $^{60}$Co treatment, this underestimation was approximately 4%. However, the dose to the upper thoracic and middle oesophagus, which lies under the trachea and the sternum, could be fairly well estimated because the overestimation of dose in air and the underestimation in bone compensate for one another [42].

2.2. Treatment plan reconstruction

The treatment plan recapitulates the irradiation beams setup used to treat the patient. It includes the characteristics of the irradiation source, its different positions during treatment, relative to the patient, and the position of any accessories used to shape the treatment beam geometry or the dose distribution. The treatment plan is usually included in the patient’s RT record. The description of the irradiation source is generally limited to the nature of the radiation types used to treat the patient and their energy. The exact series model of the treatment unit is seldom provided, although this information affects the out-of-field doses [21].

To evaluate the impact of missing data in the patient RT record on dose estimation to the heart during breast cancer irradiation, Taylor et al [30] tested three beam setup parameters: the source-to-skin distance, the field border position and the boost position. Each parameter was varied within a comprehensible range and its impact on the mean dose to the heart was estimated. Source-to-skin distance was found to have little impact on the mean heart dose. The source-to-skin distance was varied from 70 cm to 100 cm. For IMC setups, the dose difference between the highest and lowest calculated dose was about 0.2 Gy. In the case of a tangential pair irradiation treatment plan, the dose difference was about 0.1 Gy. Field border
position could vary with visual identification of landmarks. Its impact on the dose was quantified in the case of a tangential pair and direct IMC fields for a $^{60}$Co breast RT. The mean dose to the heart with the standard position was compared with the mean dose with the fields moved superiorly by 1 cm and inferiorly by 1 cm. The difference in mean dose for such movements was less than 20%. Similarly, for the boost position, a 2 cm shift was performed which led to a 100% increase in the mean dose (from 0.9 Gy to 1.8 Gy when the $^{60}$Co boost was moved 2 cm towards the heart).

In another work, Lamart et al [41] reconstructed the dose to the oesophagus from breast cancer RT. Among 414 patients, 266 had a supraclavicular anterior field. Although the exact position of the oesophagus relative to the medial border of the field was uncertain, the authors assumed that the oesophagus was at the edge of the field. Hence, if the oesophageal dose was estimated in this configuration but it was actually outside the treatment field the dose would be overestimated by a factor of three.

2.3. Characterisation of the irradiation source

Uncertainty due to the characterisation of the beams occurs when data on the dose distribution from the treatment unit are not available or when information on the beam energy or the treatment unit series model is not available. To reconstruct the dose received by patients when the original treatment unit is no longer available, medical physicists have used similar but more modern machines. Stovall et al measured the dose in an anthropomorphic phantom (Alderson-Rando, RSD, Long Beach, CA, USA) irradiated with a Philips RT250 orthovoltage machine (Rahway, NJ, USA), a typical treatment unit used in the early 1950s [1, 43], to reconstruct the dose received by 4000 paediatric patients treated with orthovoltage RT for lymphoid hyperplasia of the tonsils between 1939 and 1962. Similarly, when calculating doses, treatment-unit-specific data are needed to calibrate the algorithm. Hence, if the treatment unit is no longer available, data from a similar treatment unit may be used. However, for a similar dose delivered at the isocentre, out-of-field doses from one treatment unit to another may vary up to nine times [44]. Van der Giessen [45, 46], in the 1990s, and Joosten et al [44], in 2011, measured out-of-field doses from several treatment units. Van der Giessen tested four cobalt machines and 37 linear accelerators from seven manufacturers, while Joosten et al tested five linear accelerators from four manufacturers. Both reported substantial differences between the out-of-field doses of the evaluated machines. From the data reported in their papers, we calculated that, in the case of van der Giessen, the out-of-field doses could vary up to four times from one treatment unit to another. Joosten et al reported up to a nine-fold difference amongst the tested accelerators.

The authors mainly attributed these differences to the architecture of the beam-defining collimator because the geometry used could differ from one manufacturer to another: some had only two levels of collimation along the X- and Y-axis, some had a multileaf collimator (MLC) in addition, some had the X-jaw collimator at the bottom and others the Y-jaw collimator, etc.

For a radiation field of 20 cm $\times$ 20 cm, at 10 cm water depth and 30 cm from the beam axis, van der Giessen reported a mean dose of 0.225% of the maximum dose and a coefficient of variation of 21% for all tested accelerators. Similarly, for a 10 cm $\times$ 10 cm field at 50 cm from the beam axis, the mean dose was 0.044% and the coefficient of variation 33% [46]. The coefficient of variation may be of interest for determining the uncertainty of out-of-field dose evaluation caused by using a surrogate beam of similar quality.

Taylor et al [30] tested the impact of the use of two different makes of linear accelerators, with the same photon beam (6 MV), on the reconstructed mean dose to the heart during breast
cancer RT. At most, they found that the dose could vary between 0.8 Gy and 1.4 Gy. Although they concluded that the use of different RT treatment units had little impact on the mean heart dose, it still represents a 75% increase in dose.

2.4. Estimation of the dose at the point or volume of interest

The dose delivered to the patient, according to the treatment plan, can either be measured or computed. Measurements are performed using a physical phantom whereas calculations require a representative patient RT-planning CT or a virtual phantom.

Thermoluminescent dosimeters (TLDs) are typically used to measure the doses inside an anthropomorphic phantom \[15, 18, 20–24, 47–50\]. TLDs are inserted into holes readily available in most phantoms. The holes can be distributed along a 3D grid or according to average organ positions. Other detectors that may be used include ionising chambers or radiochromic films \[19, 51–53\].

With regard to calculations, the dose delivered according to a treatment plan is usually estimated using a clinical treatment planning system (TPS). TPSs are commercial software which include several algorithms to select the best beam configuration to deliver the prescribed dose to the target volume while limiting the dose to organs at risk \[54\]. As different algorithms are included in different TPSs, the dose evaluation can vary from one TPS to another depending on the algorithm and the treatment plan. The main algorithms available in present clinical TPSs \[55\] are measurement-based \[56, 57\] and model-based algorithms \[58–64\]. However, TPSs are not designed to evaluate the dose in the regions of the patient’s body outside the treatment beams \[44, 65–75\]. This has led medical physicists to develop in-house algorithms for estimating out-of-field doses. Three main categories of out-of-field dose algorithms exist: measurement-based algorithms \[6, 14–16, 72, 76, 77\], multi-source algorithms \[17, 78\] and Monte Carlo (MC) algorithms \[44, 79, 80\].

As part of an epidemiological study, Stovall et al \[15\] developed a systematic method for determining tissue doses in about 20,000 patients who were treated for cancer of the uterine cervix in many institutions in the United States, Canada and Europe from 1916 to 1975. This work was a remarkable dose reconstruction effort related to second cancer studies. The authors measured and calculated doses from external-beam radiation therapy involving several treatment units. Measurements were made in an anthropomorphic female phantom. Calculations used either MC or measurement-based models to analyse the contribution of leakage and scatter radiation. Later on, Stovall et al \[81\] reconstructed the gonadal dose for childhood cancer patients as part of a multi-institutional effort to study the genetic effects of radiation therapy using previously developed methods \[15\]. Another dose reconstruction project was performed by Stevens et al \[82\] for children who underwent prophylactic cranial conventional radiation therapy. Doses were determined using both anthropomorphic and in vivo measurements. These dosimetry data were then used to improve risk models of thyroid complications in children undergoing similar treatments.

More recently, Zhang et al proposed a method involving the combination of a TPS together with measurements \[83\]. The TPS (Eclipse version 8.9, Varian Medical Systems, Palo Alto, CA, USA) was used for in-field dose evaluations only, whereas measurements were used for out-of-field dose evaluations. Thus, the authors were confronted with the need to define the limits between the in-field and out-of-field regions \[84\]. In a previous study, Howell et al \[67\] reported that the absorbed dose values for 6 MV photon beams calculated with this TPS were accurate above the 5% isodose level. Hence, Zhang et al choose to use the TPS only if an organ was within the 5% isodose, and the measurements only if the organ was
entirely outside the 5% isodose. If the organ was partially inside the 5% isodose, the dose was estimated using both techniques with a volume-weighting approach.

In medical physics, standard deviation is typically 1.5% (1 sigma) of the measured dose for ionisation chamber measurements inside the treatment beam [85]. This is due to uncertainties during the measurement process or uncertainty of the different factors (e.g. the calibration factor of the detector or the quality correction factor) when converting signal to dose [85, 86]. Measurements outside the treatment beam are subject to additional phenomena, including variation of the photon spectrum [87–93], which may influence the evaluated dose. As these spectrum variations are unknown, it is difficult to properly evaluate the out-of-field dose uncertainty. A work from Bordy et al [91] estimated the uncertainty for out-of-field measurements by optically stimulated luminescence detectors (OSL) by looking at all the sources of uncertainty and combining them in a model following the guide to the expression of uncertainty in measurement (GUM). They estimated a maximum uncertainty of 4.5% (1 sigma). They also performed measurements with several other detectors: an ionisation chamber, TLDs and radiophotoluminescent dosimeters. They stated that because a good agreement with the reference values was found, an approximated global uncertainty could be derived by quantifying the spread of the measurements for all detectors. They found the GUM method was in quite good agreement with the global method, with a maximum spread of 15%, and hypothesised that this value corresponded to a coverage factor of three.

In 2009, an IAEA pilot study quantified the range of deviations between calculated and measured doses when using a commercial TPS [94]. A total of 53 clinical test cases were studied throughout 17 hospitals, using different photon beam energies ($^{60}$Co, 4 MV, 9 MV, 10 MV, 15 MV, 18 MV and 20 MV) and 14 calculation algorithms/heterogeneity correction methods implemented in nine TPS. The algorithms implemented in the TPS were classified according to three main types: measurement-based, model-based paired with pencil kernels and model-based paired with point kernels. Ion chamber measurements (0.125 cm$^3$ PTW31010 and IC10) were achieved in an anthropomorphic phantom which followed the whole chain of RT treatment planning. Overall, average discrepancies were lower than 5% for most model-based point kernel algorithms, while larger discrepancies were found for measurement-based and model-based pencil kernel algorithms. As a general trend, deviation between measurements and calculations increased with beam energy. For high-energy x-ray beams (10–20 MV) and simple algorithms (measurement-based and model-based pencil kernels) local dose differences could be as much as 24%.

Aspradakis et al [95] reported a similar experiment for two algorithms available for the Helax-TMS TPS, one pencil kernel-based and one point kernel-based. In addition to in-field measurements, these authors compared TPS calculations with measurements in volumes with a high dose gradient, the treatment field border and the electronic build-up regions. Photon beams of 4 MV, 6 MV and 15 MV were tested under a range of clinically relevant irradiation geometries. Average discrepancies in these regions were less than 15%, while maximum local discrepancies could be up to 28%.

Regarding out-of-field dose evaluations from TPSs, many authors have reported that TPSs should not be used for out-of-field dose evaluations [44, 65–75]. Howell et al [67] quantified the discrepancies between TLD measurements and dose calculations using Eclipse TPS (version 8.6, Varian Medical Systems) and the AAA algorithm, a point kernel-based algorithm. The experiment was carried out for the case of a patient with Hodgkin lymphoma. The treatment planning was performed on an anthropomorphic phantom (ATOM, CIRS Inc., Norfolk, VA, USA) using the 6 MV photon beam of a Varian Clinac 2100. TLDs were inserted inside the phantom during irradiation and measured doses were compared with calculated doses from the TPS. Beyond 3.75 cm from the field edge, discrepancies were about
40% and could be up to 55% for the furthest points (11.25 cm in this study). In a similar comparison, Joosten et al. [44] found differences of up to 179% for points within 10 cm from the beam edge of a small field size (5 cm × 5 cm) when comparing the dose calculated with the TPS CMS XiO (version 4.60, Elekta, Crawley, UK), using a point kernel-based algorithm, and the dose measured with a 0.13 cm$^3$ ionisation chamber (IC10, Scanditronix Wellhofer GmbH, Schwarzenbruck, Germany).

To overcome this limitation, several teams have developed their own algorithms for out-of-field dose evaluation [1, 14, 16, 17, 44, 77–80, 96]. Measurement-based, MC and multi-source algorithms have been developed. The doses calculated by these models were compared with measurements outside square field irradiation beams (4 cm × 4 cm to 40 cm × 40 cm).

Kry et al. [79, 96] developed a MC model for a Varian Clinac 2100. They found that average local differences between calculations and measurements were 16% and 17% for 6 MV and 18 MV photon beams, respectively. However, maximum local differences could be up to 50% at some points far from the treatment field when the greatest distance observed in their study was 55 cm from the central beam axis. For a similar LINAC, a Varian Clinac, Bednarz and Xu [80] found average local differences between calculated and measured doses of about 14% and 16% for 6 MV and 18 MV photon beams, respectively.

Joosten et al. [70] partially reported a comparison between TLD measurements in an anthropomorphic phantom and the calculated MC dose. The phantom was irradiated with a five-field technique for prostate cancer. In the first 10 cm from the beam edge, the absolute difference between measured and MC doses was on average 11% (range of the differences [−3%, 23%]) and 21% up to 40 cm from the field edge (range of the differences [−47%, 44%]). Based on these results, the authors estimated the upper bounds of MC uncertainties for each anatomical region to be <20% for points within 10 cm from the irradiated volume, <45% for points up to 40 cm and <85% for points beyond 40 cm.

Multi-source modelling has recently been proposed for specific components of the out-of-field dose, the collimator scatter component for several treatment units and beam energies [17], and the leakage component for the 6 MV beam of a Varian Novalis Tx [78]. Benadjajoud et al. [17] fitted TLD measurements up to 70 cm from the central beam axis for square field sizes (5 cm × 5 cm to 30 cm × 30 cm) to calibrate their model. Discrepancies between calculations and measurements were found to be about 10%. In a similar manner, Vu Bezin et al. [78] found discrepancies between leakage measurements and calculations to be below 7% for points within 100 cm of the central beam axis.

More recently, Jägert and Newhauser [77] introduced an analytical method to calculate in-field and out-of-field doses for a 6 MV beam of an Elekta Synergy. Square field size (5 cm × 5 cm and 10 cm × 10 cm) measurements using a diamond detector were performed in a water tank. Root mean square discrepancies between measured and calculated values of total absorbed dose in water were less than 9.3%.

Modelling the out-of-field dose for a photon beam accelerating energy higher than 10 MeV requires the photon-neutron dose distribution to be modelled. Bednarz and Xu [80] modelled the neutron fluence from a Varian Clinac 2100C accelerator with a MC model. They compared their calculations with measurements for three field sizes (9 cm × 9 cm, 10 cm × 10 cm and 20 cm × 20 cm) and found that the percentage difference between calculations and measurements was 20% on average and was as high as 39%. They noted that uncertainty in neutron fluence measurements was about 10% [96] and concluded that 20% was a satisfactory agreement.
3. Impact of dose uncertainty on estimates of the late effects of RT

The impact of dose levels on the precision of risk estimates was discussed in the UNSCEAR Reports of 2006, annex A [97] and 2012, annex B [4]. However, the issue of the impact of uncertainties in dose estimates used in studies of the late effects of RT have been addressed by few authors [1]. There is obvious interest in understanding how dosimetry uncertainties might affect the results of dose–response analyses and how they should be accounted for to obtain appropriate inferences for cancer survivors. A comprehensive literature search was performed for published documents outlining the impact of the main sources of uncertainty in dose estimates on the risk estimates from studies of the late effects of RT.

3.1. Impact of uncertainties in exposure variables in studies of the late effects of RT

Specific epidemiological studies have been selected to illustrate the impact of the various uncertainties in radiation exposure variables on the estimated risks from RT late effects. Firstly, Lubin et al [98] examined uncertainties in an Israeli cohort study of children exposed to radiation in the treatment of tinea capitis and the effects of uncertainties on dose–response analyses of malignant and benign thyroid tumours. The authors found that dosimetric uncertainties were due to a variety of factors, including uncertainty based on studies of phantoms, random differences in the sizes of children of a given age, and random movements by children during treatment. The results confirmed that, within the limitations of our understanding, uncertainties in dosimetry have a minimal impact on estimates of excess relative risk (ERR) per cGy. Similarly, in a cohort of 4296 subjects who received orthovoltage radiation treatment for benign conditions in the head and neck area before their sixteenth birthday, Schneider et al reported that the overall uncertainty for the thyroid dose in an individual patient may be as large as 50% [99]. The major sources of uncertainty stated by these authors were patient movement during treatment and deviations in a child’s size from standard size. The authors found that risk estimates based on an average dose were similar among subjects with square radiation fields of treatment and those with rectangular fields for all end points. Conversely, Kry et al analysed uncertainties associated with conventional RT and different energies and equipment for intensity-modulated photon RT (IMRT). The uncertainty in the risk estimates involves many different contributions. Kry et al concluded that an uncertainty of 50% in the estimated dose may result in a significant difference in risk of a second cancer [100]. In a review, Xu et al [2] reported that in order to facilitate the evaluation of dose–response relationships as defined in epidemiological models, organ-specific dosimetry is needed. In fact, one of the reasons for considerable uncertainty in the current risk models is that actual incidences of second cancers in radiation therapy patients are difficult to interpret due to the lack of accurate organ-specific dosimetric information. In a recent work on the risk of stomach cancer after treatment for Hodgkin lymphoma, Morton et al [101] assessed the impact of uncertainties in the radiation dose to the stomach on the risk estimates. Indeed, stomach position was unknown for individual patients in their study and is likely to have varied over the course of RT. Stomach size, shape and location exhibit intra- and inter-individual variation. Therefore, these authors assumed that all patients had three stomach configurations: a typical stomach shape and two alternative stomach configurations that represented the typical variation expected in the population of patients in the recumbent position. The first alternative stomach configuration, which has been correlated with massive body build and higher weight, represents a higher than usual position and reduces the lateral distance between the spine and lesser curvature of the stomach. In contrast, the second alternative stomach configuration, which has been correlated with thinner body build and
lower weight, represents a lower than usual position and slightly increases the lateral distance between the spine and lesser curvature the stomach. Morton et al [101] found that uncertainties in the radiation dose to the stomach due to the unknown of stomach position size and shape led to a variation of the estimated risk of between 4.5 and 5.8 for a radiation level ≥25 Gy, yet an underestimate of risk about 22%. Furthermore, in stratified analyses, the risk can be underestimated by up to 68%.

### 3.2. Impact of model uncertainties in RT late effects studies

Additional uncertainties pertain to the choice of models for characterising risk or the temporal patterns of risk. Few contributions that provided uncertainties in risk estimates in RT late effects studies have been outlined in the literature. Previously, Kry et al [100] performed an evaluation of the current risk models (absolute risk estimates and ratio of the risk estimates). The two risk models were based primarily on the incidence of radiation-induced malignancies in Japanese individuals. They found that very large uncertainties were associated with the absolute risk model, thus making it difficult to distinguish between the risks associated with the different treatment modalities considered. The ratio of the risk between different treatment modalities was found to provide a more statistically significant comparison than the absolute risk.

Fontenot et al [102] performed a study on the propagation of errors and several sensitivity tests to estimate the uncertainty in the ratio of relative risk following passively scattered proton therapy compared with that of IMRT. The authors reported that the uncertainties of the risk projections calculated in this manner depended on many factors, including the endpoint (cancer incidence versus mortality), the type of risk (relative risk versus absolute risk), the accuracy of the therapeutic and stray radiation dose distributions, the relative biological effectiveness of the radiation for carcinogenesis, the radiation type (photons versus protons), details of the beam delivery (scanned versus scattered proton beams) and the age and gender of the patient. Similar to Kry et al [100], the authors reported that the calculated relative risk was insensitive to the largest contributors to the uncertainty.

In order to assess the uncertainty in risk prediction by empirically evaluating the uncertainties in the parameters of a widely used risk model for various organs, Nguyen et al [103] performed an extensive literature review on second cancer studies. They compared the collected data with the prediction of the model using published parameters and determined a new set of parameters by fitting the model. The parameters used in these models were determined with limited data from epidemiological studies. Risk estimations were thus associated with considerable uncertainties. Standard error propagation was subsequently applied to assess the uncertainty in the risk model. Second cancer risks were then calculated for five paediatric patients treated for cancer in the head and neck region. The authors reported uncertainties in excess of 100% of the risk for almost all organs considered. This uncertainty in the dose–response curve in cancer risk models makes it currently impractical to predict the risk for an individual external beam treatment. On the other hand, the ratio of absolute risks between two modalities is less sensitive to the uncertainties in the risk model and can provide statistically significant estimates.

### 4. Conclusions and recommendations

We have hopefully covered the major studies that have addressed uncertainties in RT dose reconstruction and studies that investigated the impact of these uncertainties on estimates of the risk of side effects. As presented in figure 1, the dose assessment step is the central part of...
the dose reconstruction process. Input data regarding the patient anatomy, the beam setup and
the irradiation source are fed into the dose assessment algorithm to yield a 3D dose matrix. At
the present time, clinical dose assessment algorithms provide acceptable accuracy inside RT
treatment fields, with an average uncertainty of about 5%. Outside the treatment beams,
specific calculation tools must be used. An increasing number of medical physics teams are
developing their own algorithms for whole-body dose estimation with acceptable uncertainty
(often under 20%) in square fields and should reach an acceptable level of accuracy for any
photon beam in the near future. Hence, the largest uncertainties arise from the data entered
into the algorithms, in particular when such data are missing. Our work allowed us to
highlight the largest uncertainties for the different inputted data. For the characterisation of
the irradiation source, out-of-field doses from different makes of linear accelerator could vary
within 30% when the same dose was delivered at the isocentre. When patient anatomy is
unknown, mean dose variability could be up to 40%. As for the treatment plan, estimation of
the dose could vary by a factor of three depending on the exact position of the field border.

Two challenges arise from our work: to better estimate the uncertainty in present studies
and to reduce the uncertainties for future studies. To proceed with the first challenge, we
suggest the following actions:

(a) Evaluate the uncertainties in custom algorithms for out-of-field dose estimation under
clinical conditions.
(b) Broaden knowledge of the out-of-field dose variability among different makes of
treatment unit beyond 30 cm from the isocentre, and the out-of-field dose variability
when substituting one beam energy for another.
(c) Evaluate the dose variability among complete treatment plan records for the same
location, time period and institution to evaluate the expected uncertainty for similar
incomplete records.
(d) Evaluate the variability of the dose distribution between a representative sample of RT
patients and their corresponding anthropomorphic phantoms.

More accurate dose estimations cannot be reached if data are missing. Hence, we need to
be rigorous with the present data collection. The computer era has undeniably enabled more
accurate data collection (treatment plans and target volume imaging). However, some of the
required data are still not being collected. Hence, we make the following suggestions to
improve the accuracy of future studies:

(a) Implement quality control for out-of-field dose estimation, similar to the clinical quality
control that already exists in RT. This will require:
   • Establishment of out-of-field dose reference values to set the usual magnitude of the
     out-of-field doses received by patients in RT.
   • Comparison of customised out-of-field dose algorithms to these reference values and
     more extensively to practical clinical cases.
(b) Systematically collect data to characterise irradiation sources outside the treatment
    beams for all currently used treatment units and other older units that are still available.
    (c) Patient imaging should not only include the anatomical structures needed to plan the
        treatment, but also the structures to which known late effects might arise after treatment.

It is hoped that these recommendations will eventually lead to a concerted effort at the
national and international levels in order to define quality criteria for dosimetry in studies of
late effects of RT. These quality criteria will be based on guidelines for dose estimations in
epidemiological studies of RT late effects, quality assurance standards that will affect team
organisation and retrospective dosimetry quality control. This will improve the accuracy and credibility of dose reconstructions in these studies.

**Acknowledgments**

The research leading to these results has received funding from PeriDoseQuality, an INSERM-Plan-Cancer research project (grant no. ASC14017LSA) and from the 7th Framework Programme of the European Commission under Health.2010.2.4. 1–7 grant agreement n° 257505 (PanCareSurFup). The authors would also like to acknowledge the Agence Nationale de la Recherche (ANR, Hope-Epi).

**Conflict of interest**

None.

**References**


[61] Ahnesjö A, Andreo P and Brahme A 1987 Calculation and application of point spread functions for treatment planning with high energy photon beams Acta Oncol. 26 49–56
[76] van der Giessen P-H 2001 Peridose, a software program to calculate the dose outside the primary beam in radiation therapy Radiother. Oncol. 58 209–13
[90] Nunn A A, Davis S D, Micka J A and DeWerd L A 2008 LiF:Mg,Ti TLD response as a function of photon energy for moderately filtered x-ray spectra in the range of 20–250 kVp relative to 60Co Med. Phys. 35 1859–69