



**HAL**  
open science

# The bootstrap method to improve statistical analysis of dosimetric data for radiotherapy outcomes

Abdulhamid Chaikh, Jean-Pierre Bresciani, Jacques Balosso

## ► To cite this version:

Abdulhamid Chaikh, Jean-Pierre Bresciani, Jacques Balosso. The bootstrap method to improve statistical analysis of dosimetric data for radiotherapy outcomes. *International Journal of Cancer Therapy and Oncology*, 2017, 5 (1), pp.512. 10.14319/ijcto.51.2 . hal-01980916

**HAL Id: hal-01980916**

<https://hal.univ-grenoble-alpes.fr/hal-01980916v1>

Submitted on 14 Jan 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# The bootstrap method to improve statistical analysis of dosimetric data for radiotherapy outcomes

Abdulhamid Chaikh<sup>1,2</sup>, Jean-Pierre Bresciani<sup>3,4</sup>, Jacques Balosso<sup>2,4</sup>

<sup>1</sup>Department of Radiation Oncology and Medical physics, University Hospital of Grenoble Alpes (CHU-GA), France

<sup>2</sup>France HADRON national research infrastructure, IPNL, Lyon, France

<sup>3</sup>University of Fribourg Department of Medicine, Switzerland

<sup>4</sup>University Grenoble-Alpes – Grenoble, France

Received September 30, 2016; Revised May 12, 2017; Accepted June 15, 2017; Published Online August 10, 2017

## Original Article

### Abstract

**Purpose:** The purpose of this study is to validate a new technique in radiotherapy, the medical physicist needs to evaluate the dosimetric benefit and the risk of toxicity before integrating it in the clinical use. **Methods:** We validate a sound decision tool based on bootstrap method to help the radio oncologist and the medical physicist to usefully analyze the dosimetric data obtained from small-sized samples, with few patients. Statistical investigation principles are presented in the framework of a clinical example based on 36 patients with 6 different cancer sites treated with radiotherapy. For each patient, two treatment plans were generated. In plan 1, the dose was calculated using Modified Batho's (MB) density correction method integrated with pencil beam convolution (PBC) as type (a) algorithm. In plan 2, the dose was calculated using Anisotropic Analytical Algorithm (AAA) as type (b) algorithm. The delivered doses in monitor units (MUs) were compared using the two plans. Then, the bootstrap method was applied to the original data set to assess the dose differences and evaluate the impact of sample size on the 95% confidence interval (95%.CI). Shapiro-Wilks and Wilcoxon signed-rank tests were used to assess the normality of the data and determine the p-value. In addition, Spearman's rank test was used to calculate the correlation coefficient between the doses calculated with both algorithms. **Results:** A significant difference was observed between AAA and MB for all tested radiation sites. Spearman's test indicated a good correlation between the doses calculated with both methods. The bootstrap simulation with 1000 random samplings can be used for small populations with  $n = 10$  and provides a true estimation. **Conclusion:** one must be cautious when implementing this method for radiotherapy: the data should be representative of the real variations of the cases and the cases should be as homogeneous as possible to avoid bias of over/under estimation of the results.

**Keywords:** Bootstrap method, Delivered dose, Radiotherapy

## 1. Introduction

The main challenge in radiation therapy is to obtain the highest probability of tumor control or cure with the least amount of morbidity and toxicity to normal surrounding tissues. Currently, numerous different machines and techniques are used to irradiate the tumors either by photons or by protons as three-dimensional radiation therapy (3DRT), intensity-modulated radiation therapy (IMRT), tomotherapy and volumetric-modulated arc therapy, etc. On the other hand, the advance in technology provides

successive generations of Treatment Planning Systems (TPS), which include more accurate dose calculation algorithms. The new advanced techniques allowed optimizing the accuracy, the security and the clinical outcome of treatments. However, implementing the advanced techniques in photon or proton radiotherapy needs two steps. The first step: the medical physicists must assess the installation of the equipment using national and international recommendations and assurance quality protocols. The second step is to check

**Corresponding author:** Abdulhamid Chaikh; Department of Radiation Oncology and Medical Physics, University Hospital of Grenoble Alpes (CHU-GA), Grenoble, France.

**Cite this article as:** Chaikh A, Bresciani J, Balosso J. The bootstrap method to improve statistical analysis of dosimetric data for radiotherapy outcomes. *Int J Cancer Ther Oncol.* 2017; 5(1):512. DOI: 10.14319/ijcto.51.2

the dosimetric outcome of the treatment with a small group of patients “n”. If the step 2 is ignored, the expected clinical outcome could be endangered. Therefore, the physicists should provide the radiation oncologist, a tool allowing him to assess any significant alteration of the outcome and to estimate the prescription modifications associated with the implementation of the new treatment procedure. Moreover, the validation of a decision tool is an important component of quality assurance for radiotherapy. Practically, the assessment of benefit / risk of a new technology in a radiotherapy department based on a small number of patients without too much time and costs investment would be welcomed. In this study, we promote the use of bootstrap simulation using small sample sizes to simulate a larger population and adequately estimate the dosimetric alterations between two different methods of treatment planning.<sup>1,2</sup> A rather simple example was carried out using monitor units (MUs) comparison to illustrate and validate the method to help clinicians to make a decision based on statistical analysis of the differences. Consequently, the example presented here does not aim at evaluating the new treatment itself or estimating a benefit / risk balance, since treatment plan quality assessment is based on dose distribution in target and organs at risks as well as dose homogeneity. The simulation is done using 1000 random samplings derived from an original small group of patients  $n = 6$ . We present a step-by-step procedure for the bootstrap simulations. The procedure is presented using real data based on two generation of dose calculation algorithms. Finally, we discuss the question of whether the medical decision in radiation oncology can be taken based on so few patients.

## 2. Methods and Materials

### 2.1. Clinical cases and data

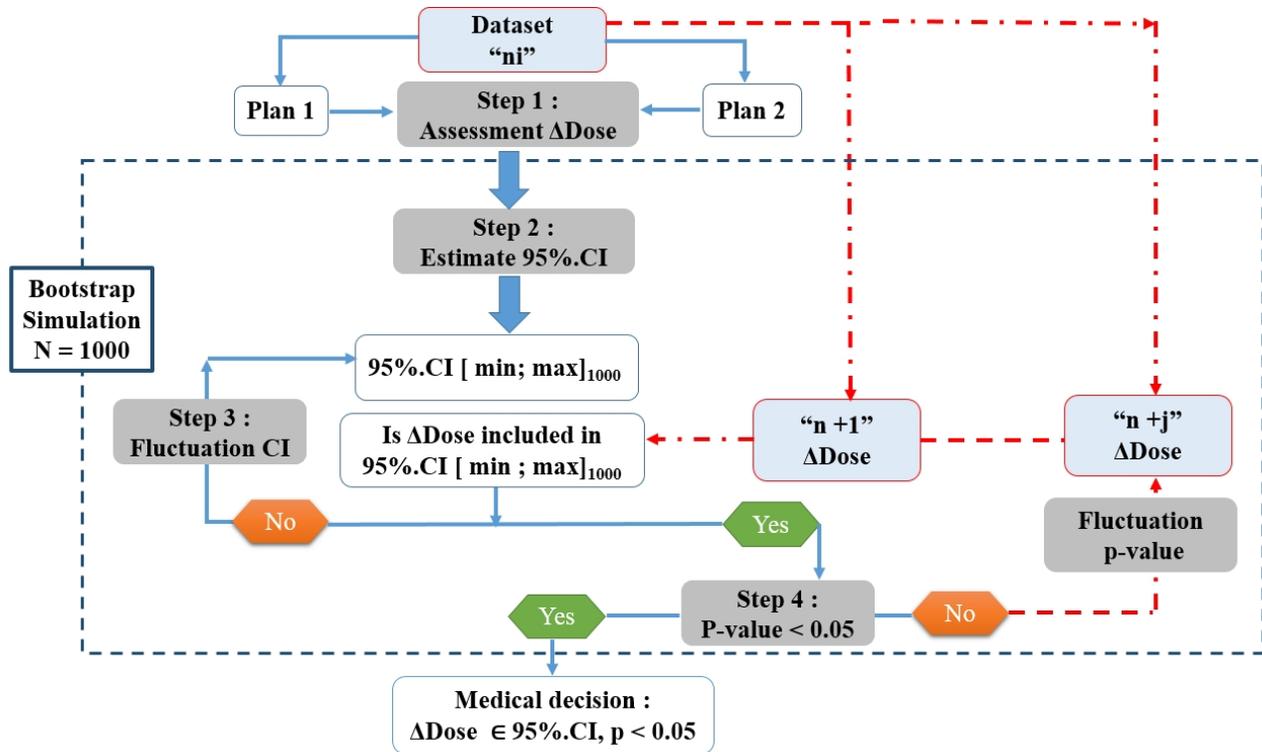
This study is based on 36 patients including 6 cancer sites. These cases were chosen to cover the full range of the different types of cancer radiotherapy, namely: lung, breast, spine, head & neck, brain and pelvis. Table 1 shows the site locations, the target volume in  $\text{cm}^3$ , the prescribed dose and the number of beams ( $m_j$ ). For each patient, two treatment plans were generated using exactly the same beam configuration. In plan 1, the dose was calculated using the Modified Batho's (MB) density correction method in combination with Pencil Beam Convolution algorithm. In plan 2, the dose was calculated using Anisotropic Analytical Algorithm (AAA). Both algorithms were integrated in Eclipse<sup>®</sup> TPS (Varian Medical Systems, Palo Alto, CA).<sup>3-8</sup> For patients treated with 3DRT, the dose in plans 1 and 2 were optimized to protect the healthy organs using respectively static filter with MB and Enhanced Dynamic Wedge (EDW) with AAA. For patients treated with IMRT, the multi leaf collimators were used in plans 1 and 2 to protect the organs at risks. The calculated dose in MUs for each plan was used to illustrate and validate the bootstrap simulation method with real data.

### 2.2. Medical decision procedures

The implementation and validation of the bootstrap simulation consists of 4 successive steps including the assessment of dose difference, estimating 95%.CI with bootstrap simulation, evaluate the fluctuation of CI and finally estimate the minimal number of cases to validate a significant difference, as shown in Figure 1.

**Table 1:** Report of tumor sites, the target volume in  $\text{cm}^3$ , the prescribed dose, n and  $m_j$  present respectively the number of patients and beams that were used for each case.

Cancer sites n = 6	Target volume [ $\text{cm}^3$ ] average $\pm$ SD	Prescribed dose [Gy] average $\pm$ SD	Techniques	Beam number $m_j$
Lung	394 $\pm$ 194	58.8 [50.8 - 66]	3DRT	34
Breast	1059 $\pm$ 248	47.2 [40 - 50.6]	3DRT	38
Spine	465.4 $\pm$ 221.6	10 [8 - 20]	3DRT	19
Head & neck	228.2 $\pm$ 135.9	56.9 [44.0 - 69.9]	IMRT	34
Brain	318.2 $\pm$ 339.1	57 [54 - 66]	IMRT	30
Pelvis	276.7 $\pm$ 249.3	65.3 [52.7 - 76]	IMRT	42



**Figure 1:** Comprehensive medical decision procedures based on bootstrap simulation to evaluate a difference between two treatment plans and make a decision when differences are significant.

**2.3. Assessment of dose difference and statistical analysis**

Initially, the dose difference ( $\Delta Dose$ ) in % was calculated between plan 1 and plan 2 using the following formula:

$$\Delta Dose \% = (D_{AAA} - D_{MB}) \times 100 / D_{AAA} \quad (1)$$

Then the 95% Confidence Interval (95%.CI) was calculated for each cancer sites with sample size  $n = 6$  patients:

$$CI = \mu \pm 1.96 \times \sigma \quad (2)$$

Where:  $\mu$  is the average deviation and  $\sigma$  is standard deviations

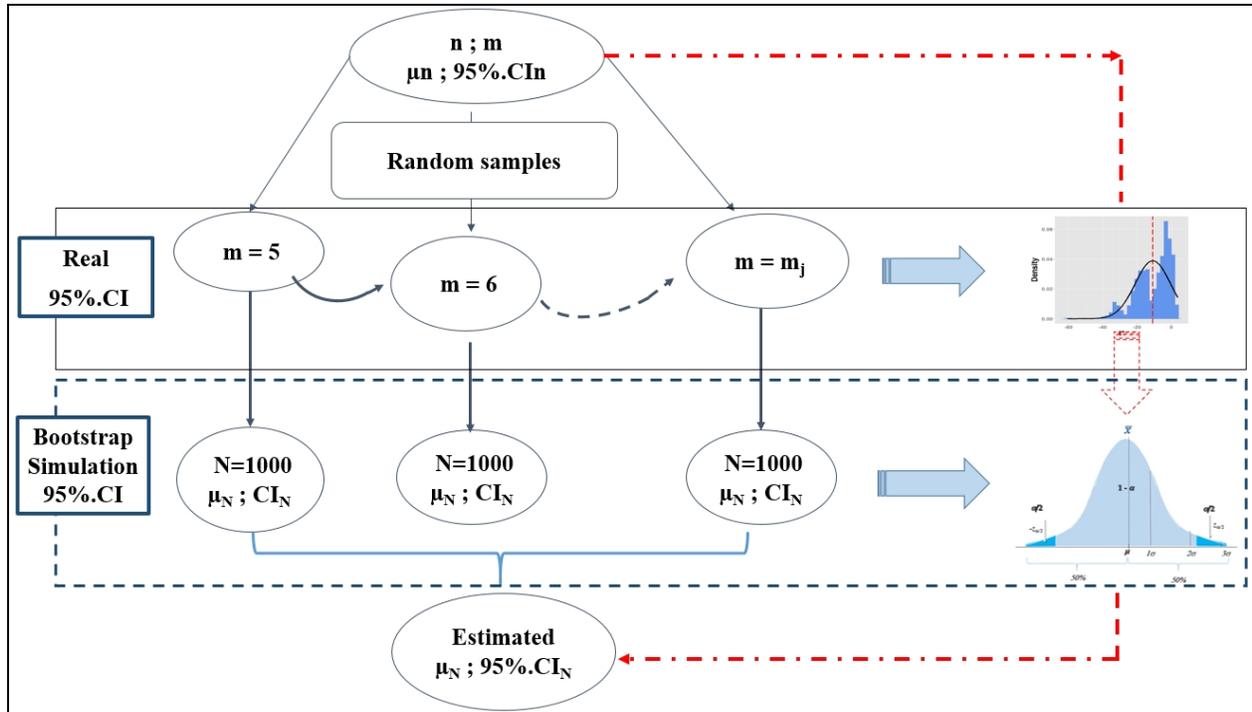
The factor 1.96 was used assuming that the data are normally distributed. Shapiro-Wilks test was used to check the normality of the data from dose difference and computing the statistic value (W) and p-value. Thus,  $p < 0.05$  means that the data are not normally distributed and that the null hypothesis  $H_0$  can be rejected; and  $p > 0.05$  means that the data are normally distributed and that the null hypothesis  $H_0$  cannot be rejected. In addition, the correlation between the doses calculated by the two algorithms was assessed using Spearman's rank test.<sup>9,10</sup>

**2.4. Data expansion to estimate the 95% confidence interval (95%.CI)**

the bootstrap method uses randomly chosen samples, iteratively drawn with replacement from the original data set, i.e., each value can be drawn several times in the same sample.<sup>1</sup> The basic idea is to artificially expand the sampling from a limited body of data in order to increase the available information and to make a better estimate of the statistical parameters of the represented population, for instance a 95% confidence interval of a given parameter, as shown in Figure 2. In this study, the CI was firstly evaluated using equation 1. Then the raw 95%.CI was simulated for each cancer site by varying the beam number from a minimum of 5 to  $m_j$ . For every sample size,  $m$ , 1000 random bootstrapped samples were drawn. Specifically, for the first round,  $m = 5$  beams were selected from the data we wanted to compare. Then for the second round,  $m = 6$  beams were selected, and so on up to  $m_j$ .

**2.5. Analysis of 95%.CI residual fluctuations**

The purpose is to assess if the 95%.CI obtained at the end the first step remains "stable" for any additional round of bootstrap for  $n+1$  to  $n+j$ , as shown in Figures 1 and 2. In fact, one can expect a small variability of the results, when new data are introduced, due to the variability of physical characteristics and anatomy of each patient. The impact has to be integrated in the results in order to readjust or to constrain the CI if the observed results with the new case alter the original CI.



**Figure 2:** Principle of bootstrap simulation with successive rounds of data expansion according to the increase number of drawn data in the sampling from  $m = 5$  to  $m_j$ .

### 2.6. Estimate the minimal number of cases to validate a significant difference

After having collected the data and computed their 95%.CI, we wanted to know what could be the smallest number of cases to compute to demonstrate a significant difference. Obviously, there is no way to guess that before having a certain number of cases. Nevertheless, when having the  $n$  cases one can assess and evaluate the  $p$ -value. To solve this, we propose the following approach. For every  $m$ , the mean  $p$ -value across the 1000 random samples was computed using Wilcoxon signed-rank test. Then the  $p$ -values as function of each sample size was plotted to show the minimum sample size needed to have a significant difference.

## 3. Results

### 3.1. Assessment of dose difference

Table 2 summarizes the dosimetric and statistical results for MUs for each cancer site. This shows clearly that for lung, breast and spine, the MUs calculated with AAA method with EDG was lower than that calculated with MB method. The results of the Wilcoxon test showed that there was a significant difference between plan 1 with MB and plan 2 with AAA for all site apart from brain. The significant difference for MUs is due to the change of filter type when using AAA with EDW. The data showed a strong correlation between the two methods with  $r > 0.85$  for all sites. Table 3 presents the results of the Shapiro-Wilk test, as well as skewness for dose difference. The Shapiro-Wilk test shows a

significant deviation from normality. Figure 3 shows the observed dose difference compared with expected data for normal distribution using Shapiro-Wilk test. Visual inspection of Figure 3 confirms the results presented in Table 3.

### 3.2. Bootstrap simulations

#### 3.2.1. Estimate the confidence interval

Table 4 shows the statistical results for each cancer site with  $n = 6$  patients using bootstrap of 1000 replicates, simulated by varying sample size from  $m = 5$  to  $m_j$ . Figure 4 shows bootstrap distributions of dose difference based on 1000 replications, for sample sizes of  $m = 5, 10, 15, 20$  and 38 for breast cancer. It is clear that using higher sample sizes data distribution is closer to normality.

#### 3.2.2. Fluctuation of confidence interval

Table 5 shows the statistical results for each cancer site with the  $n = 5$  and  $n = 6$  using bootstrap of 1000 replicates. Figure 5 shows bootstrap distributions of dose difference and cumulated average difference based on 1000 replications, for  $n = 5$  patients compared to  $n = 6$  patients for lung cancer. It can be seen that the average difference at probability of 50 % was -18 % using  $n = 5$  or  $n = 6$ . However, the lower and upper confidence interval was changed. The red-circled landmarks in Figure 6 indicate the fluctuation interval presenting lower and upper limits of dose difference.

**3.2.3. P-values**

We observed a significant difference between MB and AAA for all cancer sites with  $p < 0.05$ , a part from brain. Figure 6 shows the computed mean p-values for each sample size for all cancer sites. It can be seen that with  $m$  close to 10 beams, we can observe a significant difference between MB and AAA for lung and spine. However, with  $m > 25$  we can also observe a significant

difference for head & neck and pelvis, but we cannot conclude that there is a significant difference for brain even with  $m = 30$ . This is due to the lower dose difference, which was  $< 2\%$ . The results obtained from Figure 6 with bootstrap simulation confirm the observed results in table 2.

**Table 2:** The dosimetric and statistical results for  $\Delta MU$ s for each cancer site.

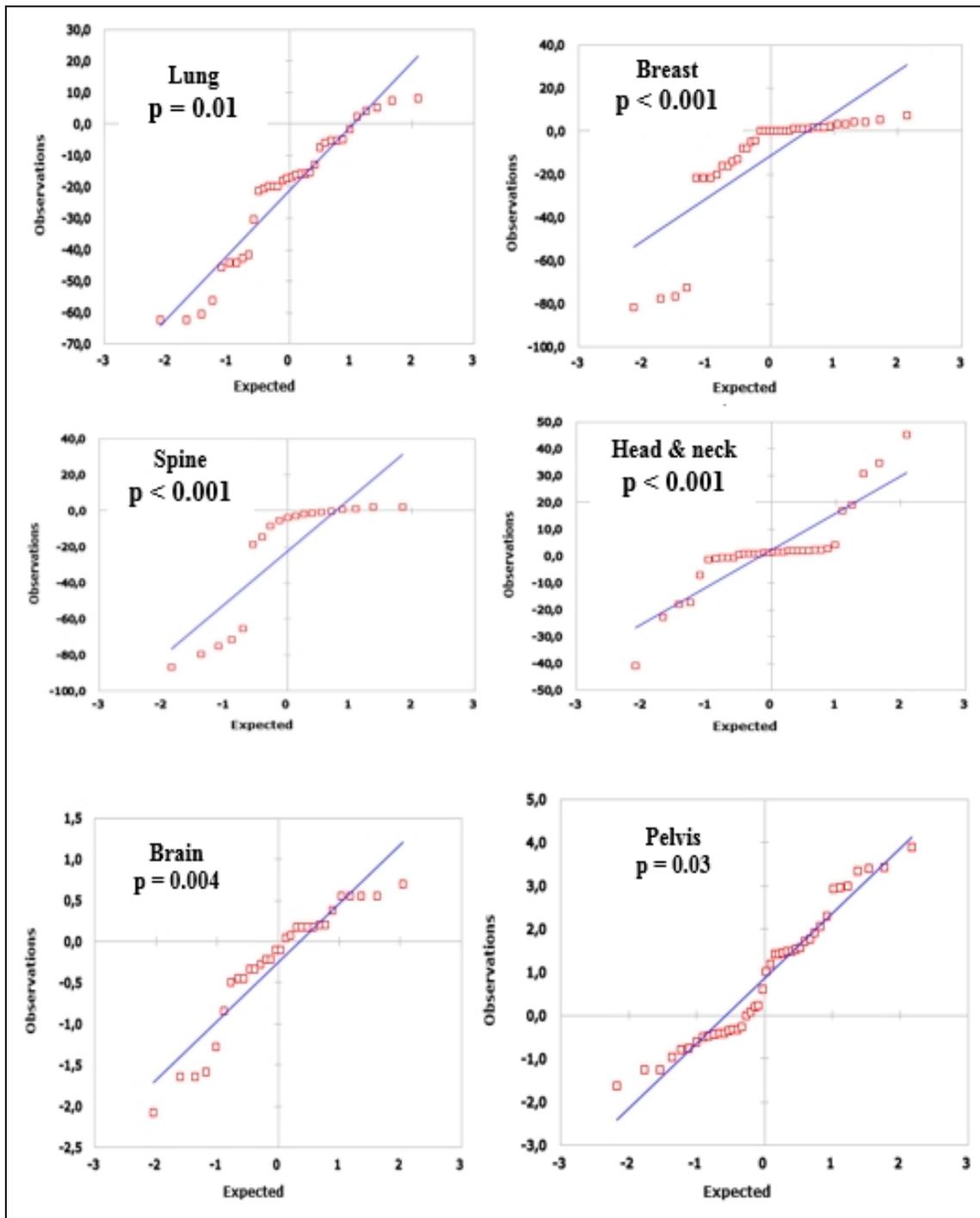
Cancer sites	$\Delta MU$ s $\mu \pm \sigma$	95%.CI Equation 2	r-value	p-value
Lung	-21.3 $\pm$ 20.7	[- 61.8 ; 19.3]	0.90	< 0.001
Breast	-11.5 $\pm$ 33.2	[- 59.1 ; 36.0]	0.92	0.002
Spine	-22.6 $\pm$ 36.0	[-87.7 ; 42.5]	0.85	0.003
Head & neck	2.0 $\pm$ 15.0	[-27.8 ; 31.8]	0.96	0.01
Brain	-0.2 $\pm$ 0.7	[- 1.7 ; 1.2]	0.99	0.3
Pelvis	0.84 $\pm$ 1.5	[-2.1 ; 3.7]	0.99	0.001

**Table 3:** The results of the Shapiro-Wilk test, as well as skewness for dose difference. No means that the data are not normal distributions.

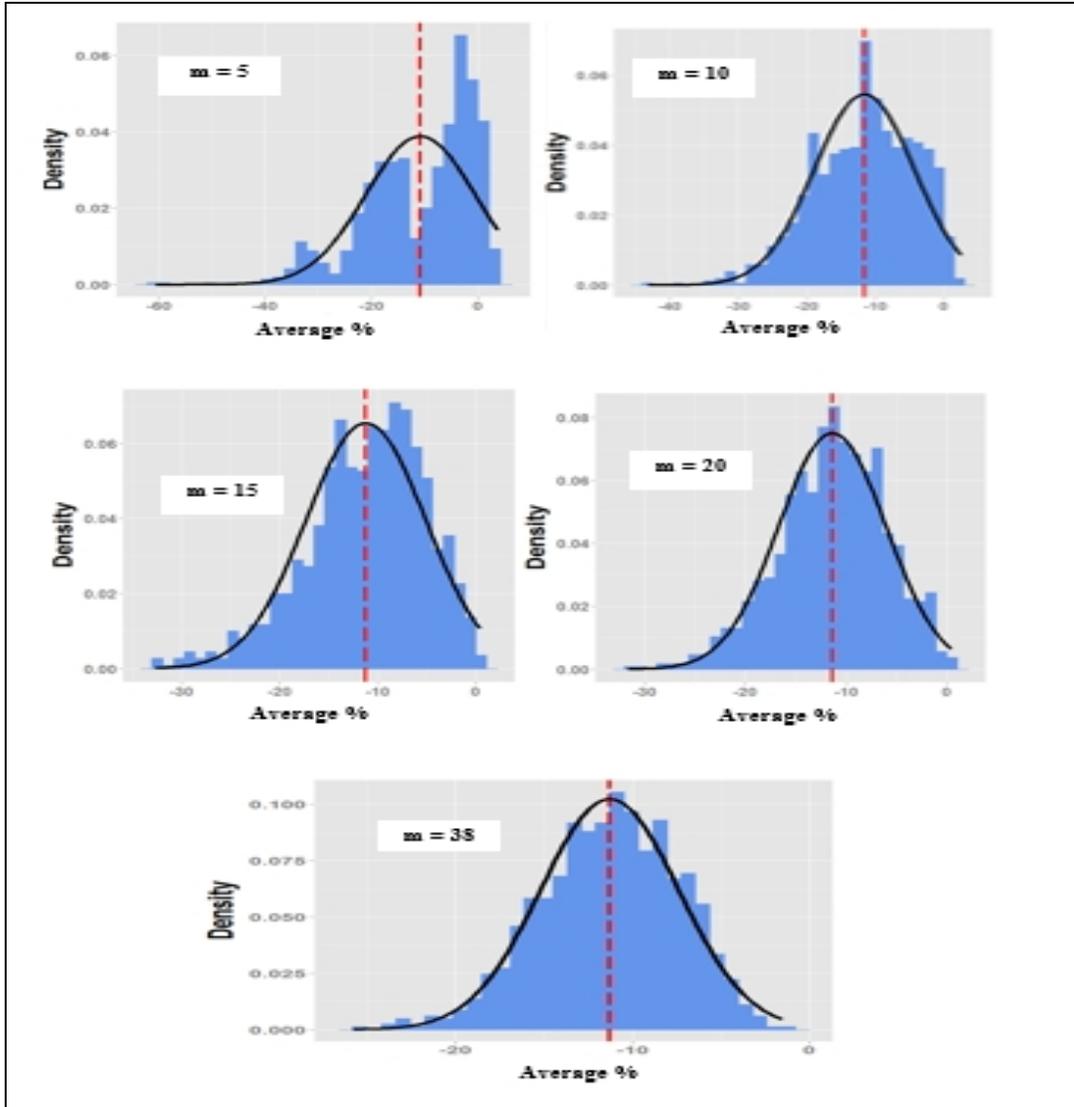
Cancer sites	W-statistic	p-value	Normality	Skewness
Lung	0.9	0.01	No	-0.6
Breast	0.6	< 0.001	No	-2.2
Spine	0.7	< 0.001	No	-1.1
Head & neck	0.8	< 0.001	No	0.3
Brain	0.9	0.004	No	-1.0
Pelvis	0.9	0.03	No	0.3
Lung	0.9	0.01	No	-0.6

**Table 4:** The statistical results of confidence intervals for each cancer site. The data were derived from 6 patients and bootstrap of 1000 replicates simulated by varying sample size from  $m = 5$  to  $m_j$ .

Cancer sites	95%.CI $m = 5$	95%.CI $m = 10$	95%.CI $m = 15$	95%.CI $m = 20$	95%.CI $m = m_j$
Lung	-39.7 ; -4.3	-34.3 ; -8.5	-31.1 ; -11.3	-30.5 ; -12.3	-28.0 ; -14.7
Breast	-30.9 ; 9.3	-25.9 ; 2.7	-23.2 ; 0.7	-21.7 ; -9.3	-18.9 ; -3.6
Spine	-51.4 ; 5.9	-43.2 ; -1.9	-39.6 ; -6.3	NA	-37.3 ; -8.2
Head & neck	-10.7 ; 15.7	-6.8 ; 11.2	-5.5 ; 9.5	-4.9 ; 8.9	-2.8 ; 6.8
Brain	-9.2 ; 0.4	-0.7 ; 0.2	-0.6 ; 0.1	-0.6 ; 0.06	-0.5 ; 0.01
Pelvis	-0.5 ; 2.1	-0.1 ; 1.2	0.06 ; 1.6	0.2 ; 1.5	0.4 ; 1.3



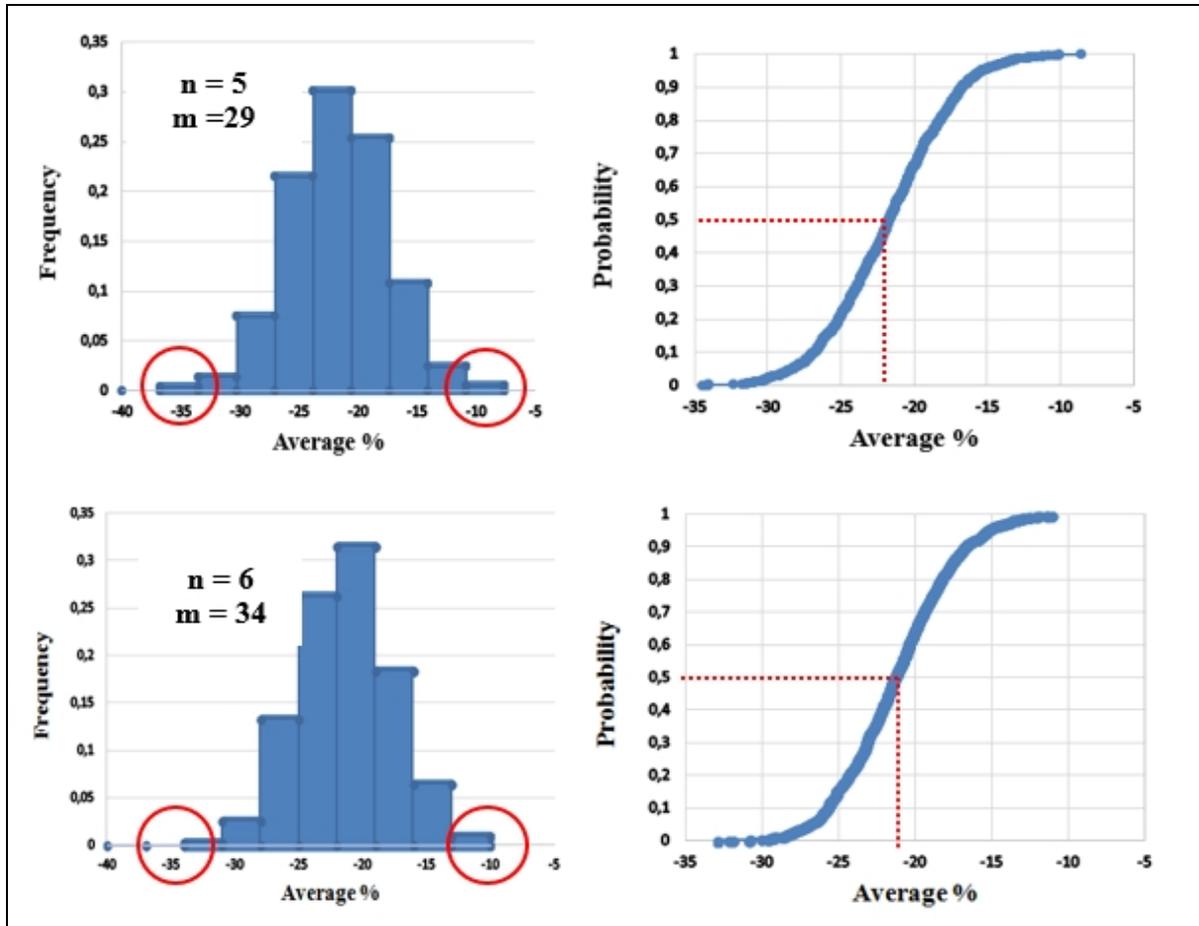
**Figure 3:** Representation of the observed dose differences compared with expected z-score obtained according to the normal distribution using Shapiro-Wilk test.



**Figure 4:** Bootstrap distributions of dose difference based on 1000 replications, for sample sizes of  $m = 5, 10, 15, 20$  and  $38$  for breast cancer.

**Table 5:** The statistical results of 95% confidence intervals for each cancer site. The data were derived with bootstrap of 1000 replicates from  $n = 5$  and  $n = 6$  patients.

Cancer sites	95%.CI $n = 5$	95%.CI $n = 6$
Lung	-30.1; -13.6	-28.0; -14.7
Breast	-22.2 ; -4.7	-18.9 ; -3.6
Spine	-39.1 ; -5.2	-37.3 ; -8.2
Head & neck	-3.6 ; 8.7	-2.8 ; 6.8
Brain	-0.5 ; 0.05	-0.5 ; 0.01
Pelvis	0.4 ; 3.3	0.4 ; 1.3



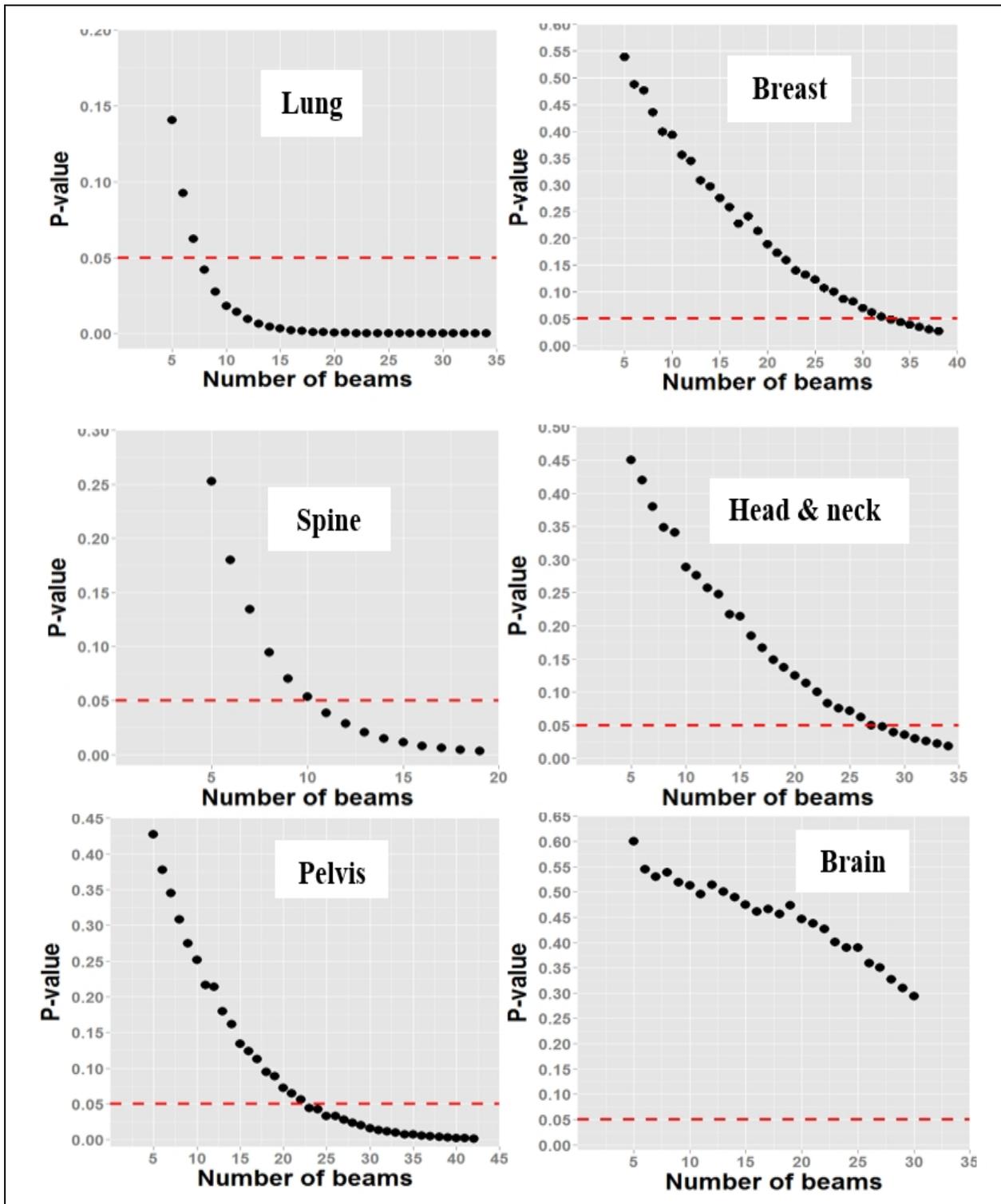
**Figure 5:** Bootstrap distributions of dose difference based on 1000 replications, for  $n = 5$  compared to  $n = 6$  patients for lung cancer. The red-circled landmarks indicate the fluctuation interval presenting lower and upper limits of dose difference.

## 4. Discussion

The bootstrap is a computer-based method proposed by Efron *et al*, 1993 to simulate the sample distribution around a meaningful statistic value (e.g., mean, median, variance, correlation coefficient) by generating multiple random samples with replacement.<sup>1,2</sup> Recently, this method was implemented in radiation oncology to estimate toxicity, setup errors, organs motions or cost, etc.<sup>11-13</sup> In this study, bootstrap simulation was used with two aims; i) to estimate the 95%.CI for small sample size; ii) to estimate the minimum sample size that would have been necessary to observe a significant difference between two algorithms MB and AAA. It is obvious that the use of EDW filter with AAA calculation for lung, breast and spine with 3DRT had a significant impact on MUs when using the same beam configuration with another filter and a different calculation algorithm. Nevertheless, this is presently used just as an example to generate a set of differential data to demonstrate a statistical procedure to support a medical decision. However, in routine activity rather large differences are naturally observed depending on cancer site, anatomy and beam orientations.

### 4.1. Relationship between data normality and 95%.CI

Regarding normality test, we observed that the data were non-normally distributed for all cancer sites. Thus, the 95%.CI based on equation 1, assuming a normal distribution, was overestimated for all cancer sites, as shown in Table 2. However, when simulating the data with a random sample of 1000 observations, the distribution mean appears to be smaller compared to the original estimation with  $n = 6$ , as is the estimated 95%.CI. This is due to the fact that the original body of data was too small to have a normal distribution. This explains why the 95%.CI values presented in Table 2 do not estimate the true interval limits of dose differences. Therefore, the bootstrap simulation allows for quantifying the difference using 1000 random samples, while avoiding the observed over/underestimation of dose difference. Before estimating the 95%.CI, one should test whether the data are normally distributed to avoid the wrong conclusion. For example, assuming a normal distribution for data, as shown in Table 2, the 95%.CI for lung, breast and spine spans zero. One could therefore conclude that there is no significant difference. However, using normality test the p-value is  $< 0.05$  demonstrating that data are not normally distributed.



**Figure 6:** Representation, using the patient’s data of the present study, of p-values estimated by bootstrap procedure, indicating the average p-value for each sample-sizes going from 5 to  $m_j$ . The red dashed line corresponds to a significance threshold of 0.05.

#### 4.2. Validation of bootstrap simulation method

We observed significant differences between MB and AAA for lung and spine. To assess whether bootstrap simulation allows for predicting significant differences for small sample size, a simulation was performed to calculate p-value as a function of sample size. The mean p-value calculated with 1000 random samples confirm the observed results, as shown in Figure 7. It also can be seen from Tables 4 and 5 that for  $n = 5$  or  $n = 6$ , the 95%.CI values were closer to the "true" CI based on all fields  $m_j$  for each cancer site. Thus, the results deriving from 1000 random provide evidence that the simulation estimates a valid 95%.CI.

The use of bootstrap method can prevent the over or under estimation of 95% CI as mentioned above. The comparison of 95%.CI estimated from  $m = 5$  to  $m_j$  showed a considerable difference, as expected, due to the anatomical variability between patients, especially for lung, breast and spine cancer sites.

The results from this study confirm that a careful analysis should be taken when using bootstrap simulation in radiotherapy especially for large tissue heterogeneity like for lung and breast. The chest cancer provides heterogeneous data due to variation of lung density from patient to patient. The realistic example taken in this study, which is based on the integration of more advanced algorithm AAA than the former algorithm MB, shows the complexity and the real difficulty when integrating advanced technology in radiotherapy. On the other hand, this also shows that the physicists and radiation oncologists should be cautious when integrating new technologies. However, the internal validation of estimated 95%.CI could detect the erroneous predictions of dose difference between both algorithms. This is a very important step that demonstrates the advantage of bootstrap simulation to cumulate the observed results and readjust the CI. This means that the use of new data derived from new patients, not included in the initial set of data, are able to probe the obtained CI, to check this result and to feed a new cycle of calculation of CI to more accurately predict dosimetric metrics in radiotherapy.

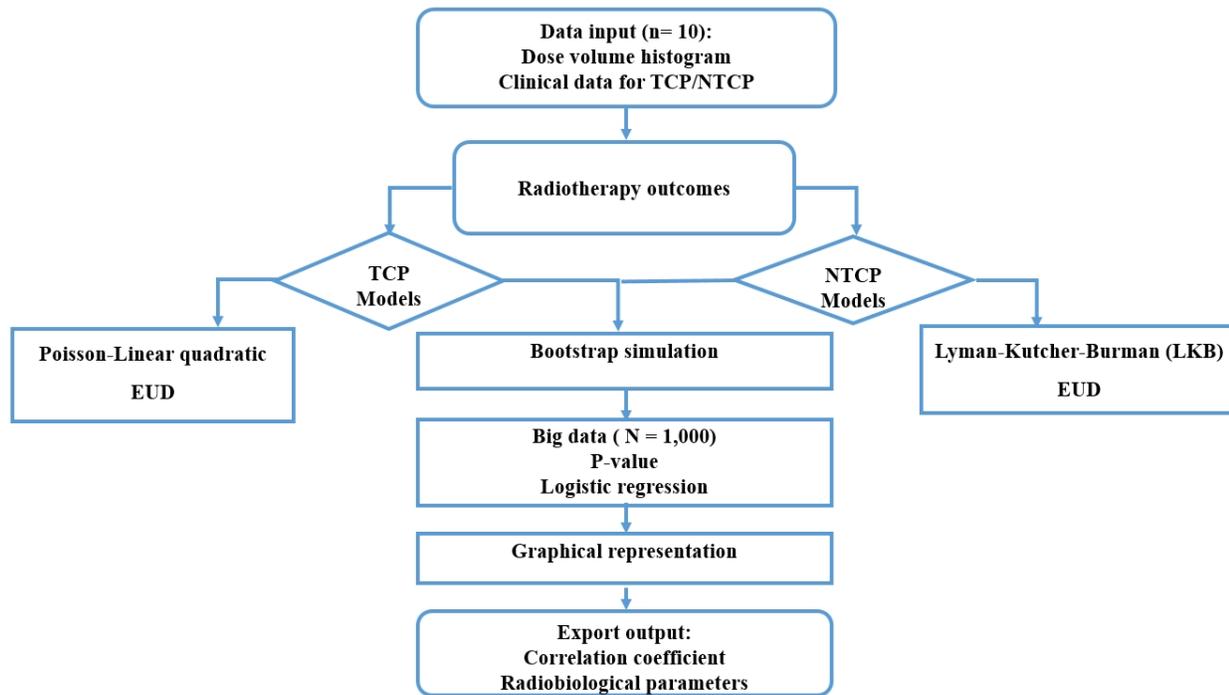
#### 4.3. Assessing the radiotherapy outcomes using bootstrap analysis

The bootstrap simulation method can be used to provide a statistical analysis of the uncertainty in the estimated dose response relation.<sup>11,14</sup> More recently Chaikh, *et al.* 2016, proposed the use of bootstrap simulation to estimate the correlation between normal tissue complication probability (NTCP) and physical lung density. They showed that the bootstrap simulation with 1000 random samplings may have under/overestimate the correlation using small data from dose volume histograms (DVH). However, bootstrap simulation can be used to re-estimate the value of radiobiological parameters setting "clinical data" for each

radiobiological model, such as Lyman-Kutcher and Burman (LKB), equivalent uniform dose (EUD), etc. The more recent study showed that the initial radiobiological parameters can introduce a considerable over or under estimation to NTCP or tumor control probability (TCP) using the advanced algorithm models, such as AAA and Acuros XB (AXB) modes: dose-to-water or dose-to-medium. In addition, a shift for the radiobiological parameters has been proposed.<sup>15</sup> The new radiobiological parameters with uncertainties can be evaluated and presented as a 95%.CI using bootstrap simulation method. In this case, the data as input includes DVH and initial radiobiological parameters for NTCP and TCP, as shown in Figure 7.

#### 4.4. Advantage and limits of bootstrap simulation in radiotherapy

The challenge is the small number of patient ( $n = 10$ ) available to produce robust 95%.CI able to sustain a decision. The bootstrap procedure provides a solution for this. By enlarging the number of data, one can see that the limits of the CI are altered as shown in table 4. When the zero is excluded of the 95% CI, one can conclude that a significant difference exists, thus making possible truly motivated medical decision. According to the type of data and the clinical situation, the possibility to reach a significant difference will need different number of cases as shown by the decreasing of p-value according to the case number as shown by Figure 6. This number is rather small when differences are large as for lung and conclusion can be drawn with a small number of cases. Of course, it is the contrary for small difference as for pelvis and  $p < 0.05$  is even not reached when probably no differences are existing as for brain. The bootstrap simulation has certain advantages and limits. First, this method can be used for small sample size. Moreover, using a large sample size, the bootstrap method should provide an even better estimation. Second, this method can be rapidly used in radiotherapy. If the simulation is properly implemented, it provides more accurate statistical values and estimates all dosimetric parameters with small cohorts of patients. Third, the simulation is fast and needs a minimum of assumptions and there are no major requirements. Fourth, the bootstrap can be used for parametric and non-parametric tests. Fifth, it can be used retrospectively to estimate the needed sample size to observe a significant difference, as presented in this study. However, one of the disadvantages of this method is that if the data does not represent the real observations, it will over/under estimate the results. In this case, one solution that was proposed in this study is to cumulate the data and readjust the confidence interval. However, a sample size larger than 10 is needed to provide a good estimation, especially if the data are heterogeneous and not normally distributed.



**Figure 7:** Estimation of clinical data for radiobiological models using bootstrap simulation method.

## 5. Conclusion

In this study, we illustrate a bootstrap analysis to estimate the 95%.CI for dose difference from paired observations. To use this method in radiotherapy, an example was used by comparing the delivered dose in MUs calculated with two dose calculation algorithms MB as type (a) and AAA as type (b). The bootstrap simulation can be used to generate big data from small number of DVH, since to validate a radiobiological model predicting tumor control and toxicity, one need a big data and too much time. Using this method, one is able to simulate the statistical values as mean, variance, correlation, confidence interval, etc., with 95% confidence.

## Conflict of Interest

The authors declare that they have no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

## Acknowledgement

This research is contributing to the Proton Share project, and is partially supported by the French research funding agency, Agence Nationale de la Recherche, in the frame of the "Investments for the Future" under the reference: France HADRON, ANR-11-INBS-0007.

## References

1. Efron B, Tibshirani RJ. An introduction to the bootstrap. Chapman & Hall, New York. 1993.
2. Efron B. Bootstrap methods: Another look at the jackknife. *Ann Statist.* 1979;7:1-26.
3. Task Group Number 65. Radiation Therapy Committee of the American Association of Physicists in Medicine. Tissue inhomogeneity corrections for MV photon beams. *Med Phys.* 2004.
4. Ahnesjö A, Aspradakis MM. Dose calculations for external photon beams in radiotherapy. *Phys Med Biol.* 1999;44 :99-155.
5. Batho HF. Lung corrections in cobalt 60 beam therapy. *J Can Assoc Radiol.* 1964;15:79-83.
6. El-Khatib E, Battista JJ. Improved lung dose calculation using tissue-maximum ratios in the Batho correction. *Med Phys.* 1984;11(3):279-86.
7. Rana S. Clinical dosimetric impact of Acuros XB and analytical anisotropic algorithm (AAA) on real lung cancer treatment plans: review. *Int J Cancer Ther Oncol.* 2014;2:02019.
8. Ojala J. The accuracy of the Acuros XB algorithm in external beam radiotherapy – a

- comprehensive review. [Int J Cancer Ther Oncol. 2014;2:020417.](#)
9. Chaikh A, Giraud JY, Perrin E, *et al.* The choice of statistical methods for comparisons of dosimetric data in radiotherapy. [Radiat Oncol. 2014;9:205.](#)
  10. The R project for statistical computing.
  11. Minna W. Assessing the uncertainty in QUANTEC's dose response relation of lung and spinal cord with a bootstrap analysis. [Int J Radiat Oncol Biol Phys. 2013;87\(4\):795-801.](#)
  12. Marcella P, Stefano M, Paola F, *et al.* Analysis of inter-fraction setup errors and organ motion by daily kilovoltage cone beam computed tomography in intensity modulated radiotherapy of prostate cancer. [Radiat Oncol. 2012;7:56.](#)
  13. Hout W, Kramer G, Noordijk E, *et al.* Cost- utility analysis of short-versus long-course palliative radiotherapy in patients with Non – Small-Cell Lung Cancer. [J Natl Cancer Inst. 2006; 98\(24\):1786-94.](#)
  14. Chaikh A, Balosso J. Correlation between pneumonitis risk in radiation oncology and lung density measured with X-ray computed tomography. [Quant Imaging Med Surg. 2016;6\(4\):413-7.](#)
  15. Chaikh A, Balosso J. NTCP variability in radiotherapy of lung cancer when changing the radiobiologic models and the photon dose calculation algorithms. [J Cancer Clin Oncol. 2016;2\(1\):100108.](#)