EDITORIAL (BY INVITATION)

Uncertainty, systemic bias, and their clinical implications in Gamma Knife surgery dose planning

Bengt Karlsson[1]

In Gamma Knife® Surgery (GKS), the absorbed dose from each source of radiation to its way to the target point must be calculated. In order to do so, measurements of the head are necessary. In the infancy of Gamma Knife Surgery (GKS), X-ray images of the skull were taken, the estimated location of the target point was defined on the images, and the AP and lateral diameters of the skull were measured. The skull measurements were thereafter used to define an ellipsoid, and the time it took to give 10 Gy to the target point was calculated, based on the assumption that the ellipsoid was filled with water. The skull scaling instrument, which is used to create a mathematical equation of the head surface (hereafter skull), was introduced in the 1980s and has not changed since. The dose planning software, initially KULA and now GammaPlan®, assumes that the radiation absorption within the skull equals that of water. The target dose, as calculated in the dose planning software, is expressed in gray (Gy), and virtually all clinical experience gained and published in GKS is based on these dose calculations. Needless to say, CT (or MR) images can also be used to define the skull, and with CT images, not only the skull can be defined, but different absorptions in different tissues can also be taken into account (convolution dose algorithm calculations).

In this issue of *Acta Neurochirurgica*, Leroy et al. report the results from a single-center study, comparing two methods to calculate the skull. One is to use the skull scaling instrument, and the other is to use the information from CT images. The difference in radiation time between dose plans using the two methods was found to be in average 1.28%. The authors raise the question if the found difference should affect future treatment doses or not. I applaud the authors for raising this issue to the neurosurgical community.

In any study involving mathematics, the significance of the calculated results must be considered. Two things are needed to be taken into consideration. The first is the uncertainty of the input data. To illustrate this, let us assume that we measure the largest diameter of three tumors to be 18, 27, and 26 mm, giving an average diameter of 23.7 mm (using one decimal). However, each measurement has an uncertainty of ± 0.5 mm. If this is taken into consideration, the average diameter will be between 23.2 and 24.2 mm. Thus, the true value will be 23.7 ± 0.5 mm or $\pm 2\%$. Let us now add an uncertainty of ± 1 mm when defining the diameters. The resulting uncertainty will now be $\pm 6\%$. It would be misleading, and most likely inaccurate, to claim that the average diameter for these three tumors is 23.7 mm. Thus, the uncertainty of the input data must be taken into consideration when reporting a result.

Another mathematical rule that is frequently ignored is the significance of the number of decimals. There is a difference between 1/3 patients and 1000/3000 patients, even though the relation is 33.3% in both examples. Let us say that the numbers represent the incidence of a complication. Let us now assume that the next patient will have a complication in both series. The complication rate will increase from 33 to 50% in the first series, but it will still be 33% in the second. Thus, the use of decimals when reporting percentages in series with less than thousands of patients is misleading, suggesting an accuracy that does not exist. Most neurosurgical articles violate at least one of these two rules.

Is the above applicable to the article by Leroy et al.? Let us first look at the skull measurements. Three experienced users measured the skull independently for five consecutive patients in our center. The average difference between the treatment times and the different measurements that resulted was 1% when comparing with the shortest treatment time, similar to that found by Nakazawa et al. [1]. It is outside the scope of this editorial to analyze if the difference was a result of a systemic bias or random, but it gives an idea of the uncertainty in

✉ Bengt Karlsson
  nykuttram@yahoo.se

1 Dep. of Surgery, Div. of Neurosurgery, National University Hospital, Singapore, Singapore

everyday clinical practice. Needless to say, there is also an uncertainty when calculating the skull based on the CT images, but the magnitude of this is for me unknown.

Based on the above, can we draw any conclusions about the 1.28% difference found? One conclusion is definite. The use of two decimals is misleading. Assuming that the differences found between the skull readings in our center is representative, it is reasonable to believe that the difference lies in the interval 0–2%. The probably lower and for me unknown uncertainty of the CT-based measurements should also be taken into consideration, which may increase this interval.

Let us take this one step further and compare to the two dose calculations above to the convolution dose calculation algorithm. In the first two, the dose absorption is assumed to be equal to that of water. In the latter, the different absorptions in different tissues are also taken into account. Bone absorbs more radiation than water, and brain (which basically has the same absorption as fat) absorbs less. Furthermore, the air/tissue build up effect can be taken into consideration in the convolution dose calculation algorithm. The difference between the two first and the last is larger and more location dependent than that found between the first two [3].

One more dimension must be taken into account. Are the differences between the methods the result of a systemic bias or random? If random, they do not have any clinical relevance. If a systemic bias, then different definitions of the skull will result in different treatment times. The first question can be answered by a literature search. There are several articles published comparing different calculation and measuring alternatives [1–4]. In fact, there are systemic biases, although small, just as Leroy et al. suggest.

The above is mathematics and physics. What is the clinical relevance of these findings? Let us take an extreme example to illustrate this. Let us assume that the dose planning formula had an inherent error, resulting in an inaccurate dose calculation. Let us say that the actual dose is twice the dose calculated in the old dose plan and that this is corrected for in a new version of the dose plan. Consequently, the treatment time for any given dose will be twice as long using the old dose plan as compared with the new, or in other words, when using the old dose plan, a prescribed dose of 13 Gy actually resulted in 26 Gy being delivered. What should we do? Start to prescribe half the dose as compared with earlier? No. We would probably define a new unit, for example, Steiner (hereafter St), and define 1 St as 2 Gy and treat using 13 St in the future. Thus, the same amount of radiation will be absorbed by the target when prescribing 13 Gy in the old dose plan as compared with 26 Gy in the new or to 13 St in both dose plans.

The reason is simply that the prescription doses we use are based on the treatment results from thousands of patients, in whom a similar amount of radiation was absorbed in the target volume. We should change the prescription dose if we think that we can improve the treatment results by changing the amount of radiation absorbed, not by any changes in calculation algorithms. The only clinical relevance in this extreme example would be when comparing LINAC with GKS results; unless, we would compare the prescription doses in St.

So, what relevance does all the above have for the article by Leroy et al.? First, it can be confirmed that there is a systemic bias when using skull measurements as compared with CT images, probably in the magnitude of around 1%. As written in the article, this means different treatment times for the same prescription dose. If your intention is to give the same dose as you always have done before, then, based on the arguments above, you theoretically need to change the prescription dose so much that the treatment time become independent of the skull measurements used. However, based on the studies by Xu et al. and Nakazawa et al [1, 2], the differences are so small that they do not need to be taken into consideration.

Does this mean that the article by Leroy et al. is redundant? No. It is of importance that clinicians understand that different measurements will result in different values. This will be even more important in the future when the use of the convolution dose algorithm calculations will become more frequently used. Thanks to the article, clinicians may take the dose calculation method into consideration in the future when prescribing the dose to eccentric located targets.

References

1. Nakazawa H, Komori K, Mori Y, Hagiwara M, Shibamoto Y, Tsugawa T, Hashizume C, Kobayashi T (2014) Effect of skull contours on dose calculations in Gamma Knife Perfexion stereotactic radiosurgery. J Appl Clin Med Phys 15:4603
2. Xu A, Bhatnagar J, Bednarz G, Niranjan A, Flickinger J, Lunsford LD, Huq MS (2014) Dose differences between the three dose calculation algorithms in Leksell GammaPlan. J Appl Clin Med Phys 15:4844
3. Xu A, Bhatnagar J, Bednarz G, Niranjan A, Kondziolka D, Flickinger J, Lunsford LD, Huq MS (2015) Gamma Knife radiosurgery with CT image-based dose calculation. J Appl Clin Med Phys 16:119–129
4. Zhao Z, Wang X, Yang J (2012) SU-E-T-445: a dosimetric comparison of Gamma Knife plans with measurement based and image based skull definitions. Med Phys 39:3807