Lessons learnt from the recent EURADOS intercomparisons in computational dosimetry

Hans Rabus [a,g,*], Maria Zankl [b,g], José Maria Gómez-Ros [c,g], Carmen Villagrasa [d,g], Jonathan Eakins [e,g], Christelle Huet [d,g], Hrvoje Brkić [f,g], Rick Tanner [e,g]

[a] *Physikalisch-Technische Bundesanstalt (PTB), Abbestrasse 2-12, 10587, Berlin, Germany*
[b] *Helmholtz Zentrum München German Research Center for Environmental Health (HMGU), Neuherberg, Germany*
[c] *Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas (CIEMAT), Madrid, Spain*
[d] *Institut de Radioprotection et de Sûreté Nucléaire (IRSN), Fontenay-aux-Roses, France*
[e] *UK Health Security Agency (UKHSA), Didcot, United Kingdom*
[f] *J. J. Strossmayer University of Osijek (MEFOS), Osijek, Croatia*
[g] *European Radiation Dosimetry Group (EURADOS) e.V, Neuherberg, Germany*

A B S T R A C T

Organized by Working Group 6 "Computational Dosimetry" of the European Radiation Dosimetry Group (EURADOS), a group of intercomparison exercises was conducted in which participants were asked to solve predefined problems in computational dosimetry. The results of these comparisons were published in a series of articles in this virtual special issue of Radiation Measurements. This paper reviews the experience gained from the various exercises and highlights the resulting conclusions for future exercises, as well as regarding the state of the art and the need for development in terms of quality assurance for computational dosimetry techniques.

1. Introduction

The European Radiation Dosimetry Group (EURADOS) is an association of 80 institutions and more than 600 individual members that promotes harmonization and good practice in dosimetry (Rühm et al., 2018, 2020; Harrison et al., 2021). EURADOS has eight working groups dealing with different aspects and application areas of radiation dosimetry. EURADOS Working Group 6 on "Computational Dosimetry" has a cross-cutting character. Its main activities include the organization of comparison exercises (Tanner et al., 2004; Gualdrini et al., 2005, 2008; Siebert et al., 2006; Price et al., 2006; Broggio et al., 2012; Vrba et al., 2014, 2015; Caccia et al., 2017) and training courses (Rabus et al., 2021a) as well as studies on fundamental aspects of computational dosimetry.

Recently, several computational dosimetry exercises have been completed, the results of which are compiled in this virtual special issue of Radiation Measurements (De Saint-Hubert et al., 2021, 2022; Eakins et al., 2021; Gómez-Ros et al., 2021, 2022; Huet et al., 2022; Rabus et al., 2021b; Villagrasa et al., 2022; Zankl et al., 2021a, 2021b, 2021c). This current article is meant as a synopsis and reflection on the common issues found in those different exercises and the lessons learnt on the state of the art in applied computational dosimetry, as well as conclusions for future exercises.

2. Overview of the exercises

The exercises can be divided roughly into two classes depending on the nature of their solutions. The first class comprised six exercises on the use of ICRP computational reference phantoms (ICRP, 2009; Zankl et al., 2021b) and one on unfolding of neutron spectra from Bonner sphere measurements (Gómez-Ros et al., 2022). The former will be referred to as "voxel-phantom exercises" throughout this article, the latter as the "Bonner sphere exercise". Despite the quite different nature of the problems to be solved, these exercises had in common that they required the application of well-known methodologies and established computational tools. This allowed the organizers to establish prior reference solutions that could be used to validate the results subsequently submitted by the participants in the exercise.

The second class are exercises where no reference solutions could be established, since one of the objectives of the exercises was evaluating the possible influence of different cross-section models in the codes used by the participants. Two of these exercises were code intercomparisons,

one dealing with the calculation of microdosimetric and nanodosimetric quantities (Villagrasa et al., 2019, 2022), and the other with the effects of gold nanoparticles on dose deposition at the microscopic scale (Li et al., 2020a, 2020b; Rabus et al., 2021b, 2021c); the former is called the "uncertainty exercise" in this article and the latter the "nanoparticle exercise". The other two exercises in this class dealt with out-of-field dose calculations. One was about calculating the dose to the foetus during maternal proton therapy treatment and the other was about calculating the secondary neutron fluence. (De Saint-Hubert et al., 2021, 2022). These two exercises are referred to as "foetus dose exercises".

It is important to note that none of the exercises was intended to be a code competition. Rather, the aim was to investigate the dispersion of results when the same problem was solved by different people using different approaches and different codes or the same code with different options. The first class of exercises focused on identifying the state of the art in the application of common methods in computational dosimetry. The second class was more exploratory in nature and aimed to assess the state of the art in terms of the capabilities of codes and approaches. All classes contained tasks of different complexity, and thus different demands on the participants' skills.

The exercises were organized and run by ad-hoc teams composed of EURADOS WG 6 members. In general, the preparation of the exercises involved independent simulations by several team members, with their respective results then cross-referenced to identify potential pitfalls in the proposed exercise definitions and to check whether the tasks were solvable based on the information to be provided. For the exercises with reference solutions, the results of these test simulations were also used to establish those values, (e.g. by taking the mean), as well as gain a handle on the typical levels of uncertainty that may be considered acceptable for them.

In some exercises, templates for reporting results were also provided to the participants.

2.1. ICRP reference voxel phantom exercises

Of the voxel-phantom exercises, two involved exposure to an external point source, emitting either ^{60}Co gamma photons or 10 keV neutrons respectively (Huet et al., 2022). The task to be solved was to calculate the organ absorbed doses and the effective dose for a given exposure duration and activity of the source.

A third voxel-phantom exercise also dealt with point source geometries, but for cases of typical X-ray examinations (Huet et al., 2022). Here the task was more complex, as participants were required to determine the position of the radiation point source in relation to the phantom. In addition, the results were to be presented as conversion coefficients to organ absorbed doses, both from air kerma and kerma area product. This exercise was thus linked to a potential practical application in which the latter quantities are determined as part of the quality assurance of radiological equipment, and the conversion coefficients sought would enable an assessment of the dose absorbed by the patient during the X-ray examination.

The fourth voxel-phantom exercise considered a uniform planar source of 60 keV photons beneath the phantom (mimicking ground contamination by ^{241}Am) and required participants to calculate organ absorbed dose rates and the effective dose rate for a given area density of the emission rate (Eakins et al., 2021). In the fifth voxel-phantom exercise, a mixed radiation field of gaseous ^{16}N was considered, emitting beta and high energy gamma radiation from both inside (lung) and outside the human body (Gómez-Ros et al., 2021); the ratios of organ equivalent dose rates to activity concentration were to be determined.

The most extensive voxel-phantom exercise involved an idealized case of internal dosimetry (Zankl et al., 2021c). For the sake of simplicity, hypothetical radionuclides were considered that were uniformly distributed in specified organs and emitted monoenergetic photon or electrons. Here, absorbed fractions and specific absorbed fractions of energy in the "source" organ and in specified "target" organs

were to be determined as well as *S*-values for the resulting source and target organ combinations for two specific radionuclides.

2.2. Bonner sphere spectra unfolding exercise

The tasks were defined by the counts measured by a set of twelve Bonner spheres of different diameters and known sensitivities (as determined by the organizers with radiation transport simulations), located at a measurement point in one of four known environments: inside the bunker of a medical linac; near a radioactive source; in a simulated workplace field within a neutron calibration facility; or outside a nuclear power plant. The count rates measured by the Bonner spheres were determined by the organizers through Monte Carlo radiation transport simulations of the respective complete measurement setup for each Bonner sphere within its environment. In addition, and to recreate a realistic situation, the count rate from one of the Bonner spheres in one of the scenarios was intentionally given an incorrect value in order to test the participants' ability to detect an erroneous measurement and exclude it when applying the deconvolution procedure (Gómez-Ros et al., 2018, 2022).

2.3. Micro- and nanodosimetric uncertainty exercise

In its first phase, this exercise included a microdosimetric and a nanodosimetric intercomparison (Villagrasa et al., 2019). In the frame of the former, frequency distributions of specific energy were to be determined within a microscopic water sphere for different distributions of a low-energy electron emitter with an energy spectrum derived from the internal-conversion Auger emitter ^{125}I. In the nanodosimetry part, ionization cluster size distributions were to be determined in target spheres of different sizes located at different distances from a point source with the same energy spectrum as in the microdosimetry part. A sensitivity analysis was also performed on the variation of inelastic cross-sections and its consequences for the calculated ionization cluster size distributions (Villagrasa et al., 2022). In the second phase of the exercise (in preparation), the focus will be on a comparison of the cross-section datasets for low-energy electron transport and a consideration of their impact on the dispersion of nanodosimetric results.

2.4. Nanoparticle exercise

In the nanoparticle exercise, the dose enhancement from a gold nanoparticle, as well as the energy spectrum of electrons emitted from it, were to be determined when irradiated with two low-energy X-ray spectra. The geometry was simply a gold sphere in water irradiated with a parallel beam from a plane photon source, the cross-sectional area of which was slightly larger than that of the nanoparticle. Two different nanoparticle diameters were considered, and the dose enhancement was to be determined in spherical water shells around the nanoparticle (Li et al., 2020a, 2020b; Rabus et al., 2021b, 2021c).

2.5. Foetus dose during maternal proton therapy

This exercise consisted of two parts. The first part dealt with the effects of different calculation phantoms for pregnant women, and different code versions of MCNP, on the predicted dose to the foetus during maternal brain proton therapy. The second part dealt with the dependence of the secondary neutron spectra on the Monte Carlo radiation transport codes and nuclear models that were used, and their effects on the calculated and measured neutron doses during proton therapy (De Saint-Hubert et al., 2021, 2022).

3. Experiences from the exercises

3.1. Observations on participants' results

In general, the ensemble of participants' results that were submitted initially showed a large scatter. In the exercises for which a reference solution was available, excellent agreement within the expected statistical fluctuations was found in some cases, while others showed significantly larger deviations, which in some individual cases ranged by up to several orders of magnitude. For the tasks without a reference solution, a subset of the reported results also agreed with each other to some extent, while others deviated significantly from this group. In both classes of exercise, the occurrence of extreme outliers was not correlated with the complexity of the problem.

Some of the deviations were attributable to simple errors, such as copy-and-paste mistakes or incorrect arrangement of the results in the given template. Others resulted from misunderstanding how the final results should be normalised (e.g. normalising to the correct quantity but at a different distance from the source than was required). In the voxel-phantom exercise for the case of X-ray examinations, some participants normalised to the value of air kerma free in air at a specific distance from the source or to the entrance surface dose (which includes backscatter) instead of to air kerma free in air at the skin as was requested. In the microdosimetric and nanodosimetric intercomparisons, the normalisation to 'one decay of the electron source' was not always understood by the participants and was also sometimes difficult to implement for some Monte Carlo codes. The use of a logarithmic scale for the microdosimetric quantity (specific energy distribution) also caused problems with proper normalisation.

Many major deviations were caused by the fact that the participants' simulations deviated from the specifications in terms of geometrical dimensions or the quantities to be determined. One example of this was the "nanoparticle" exercise, where only four out of eleven participants implemented the requested geometry correctly, which consisted of a gold sphere irradiated in water by a collimated parallel photon beam of given dimensions. Another example was the voxel phantom exercise on internal dosimetry, where some participants used organ masses that included blood instead of those given in ICRP Publication 110, as was stated in the exercise definition. In some of the voxel-phantom exercises, the choice of the location of the source was also sometimes a problem due to deviation from the correct reference point (e.g. the edge of the phantom array instead of the phantom's skin).

Other causes of major deviations were that some participants were not familiar with certain concepts, such as the normalisation quantity "kerma area product" (in the voxel-phantom X-ray exercise) or effective dose; mistakes for the latter included not applying tissue weighting factors correctly, not averaging and summing over the defined set of organs, neglecting to sex-average, or neglecting to apply the correct energy-dependent radiation weighting factor for neutron exposures.

In the voxel-phantom exercises, many participants had problems applying the method recommended for bone marrow dosimetry in (ICRP, 2010). This finding stimulated writing an article to better explain this approach, which is also part of this Special Issue (Zankl et al., 2021a).

As already mentioned, participants whose results differed from the reference solution (class 1) or from the majority of other participants' solutions (class 2) were informed of this fact and asked to revise their solutions. Not all contacted participants responded to this invitation or provided the requested information on details about their simulations.

Of those participants who submitted a revised solution, some did not indicate what they had changed in their computational procedure to arrive at their revised results. This therefore does not give any additional insight into possible similar errors to be expected in future similar exercises, or hints that could have been communicated to the other participants.

In the nanoparticle exercise, where some inconsistencies became

evident after the first publication of the results (Li et al., 2020a) and required a thorough re-evaluation (Li et al., 2020b), consistency checks provided clear indications of the causes of the discrepancies for some results. Nevertheless, some of the participants concerned did not provide revised solutions (Rabus et al., 2021b). However, it must be also stressed that the majority of participants were very supportive of the re-analysis of the results and were eager to clarify the origin of the discrepancies found initially.

3.2. Issues with omitted quality assurance of results

Many of the anomalies found in the reported data could have been detected by the participants themselves, e.g. through simple plausibility checks of their results. Examples are briefly discussed in the following.

For example, a very general plausibility consideration is that if the irradiation conditions are quite homogeneous, it may be expected that all organ doses will be of broadly similar magnitudes; a single organ dose result differing by several orders of magnitude from the rest of a given participant's dataset ought therefore to be immediately apparent to them as being potentially erroneous. Similarly, if multiple energies are considered, it is unlikely that the value for a single intermediate energy will be entirely outside the range of values for all other energies.

In the Bonner sphere exercise, there were cases of reported results with negative values for the neutron fluence. These physically impossible values, as well as anomalous spectral shapes, could have been identified by simply plotting the results. Some of the reported spectra differed from the reference solutions by several orders of magnitude; such anomalies could have been easily detected if the unfolded spectra had been convolved with the given sensitivities of the Bonner spheres, to verify that the given count rate was then achieved.

In the voxel phantom exercise for internal dosimetry, a simple plausibility check would have been that the absorbed fraction for electrons and low-energy photons must be close to unity in a source organ and quite small for other organs, since these radiations have a short range in condensed matter and therefore deposit their energy close to the point of release. Moreover, some participants in this exercise reported results for absorbed fractions and specific absorbed fractions for which the ratio of these two quantities varied between different energies of the particles emitted from the (monoenergetic) source. However, since this ratio is simply the mass of the organ, it cannot depend on the energy.

In addition, for many of the tasks in the voxel phantom exercises, literature values are available for fairly similar exposure conditions that could have been used for comparison, at least as a first approximation to indicate the expected magnitudes of the results.

When dealing with voxel phantom simulation, one of the simplest checks might be to visualise the problem in order to ensure the proper positioning of the beam, though it is noted that some software packages struggle due to the sizes of these input files. If one is using any variance reduction it should also be ensured that simulations with and without application of these techniques reproduce the same results, albeit with differing statistical uncertainties.

3.3. Issues with exercise definitions

In some cases, inadequacies in exercise definitions became apparent while they were already running. In the nanoparticle exercise, for example, one of the quantities to be reported by the participants was the energy spectrum of electrons "in spherical shells" around the nanoparticle, with the radii of the bounding spherical surfaces given.

Most of the participants interpreted this physically undefined quantity as the energy distribution of the electrons entering the respective volume. However, one participant determined the energy distribution of the balance of the number of electrons traversing the surfaces of the respective volume and withdrew her results on the assumption that the observed negative frequencies indicated an error

that she could not locate.

Another problem with this part of the nanoparticle exercise was that there was no default energy binning, so participants chose very different values for the bin size, with some using logarithmic binning and others using linear binning. The large statistical variations in the results obtained with small energy bin sizes masked the variations between the different results when plotted together (Li et al., 2020a).

In the exercise on Bonner sphere spectrum unfolding, it was found during the analysis that in one of the scenarios considered, there was an interference of the Bonner sphere response due to backscattering of neutrons from a nearby concrete wall (Gómez-Ros et al., 2018).

In the voxel phantom exercises featuring ^{60}Co photons and 10 keV neutrons, the instruction given to participants for the location of the point source was to place it '100 cm from the surface of the chest', which could be interpreted differently. In response, the organizers performed small sensitivity analyses to quantify the impact from the ambiguity of this parameter, the outcomes from which were used to imply appropriate 'tolerances' that could be applied to the submitted results (Huet et al., 2022).

The voxel phantom exercise on internal dosimetry was not wisely designed in several respects. The tasks to be solved were too extensive, which also made evaluation challenging and led to delays in feedback to the participants. For electrons and low-energy photons, the source and target organs were sometimes too far apart, which led to very large statistical uncertainties even in the reference solution. Therefore, the degree of deviation between participant and master solutions could not be reliably quantified in some situations.

In the uncertainty exercise, the use of a multi-energy electron source that was similar to the ^{125}I decay but did not take into account the variability of the actual decay, complicated the understanding of the problem on one hand and, on the other hand, did not favour the analysis of the sensitivity study on the variation of the cross sections. Indeed, the use of monoenergetic electrons would have helped in both aspects.

In the foetus dose exercise, atomic numbers of the elements, mass numbers of the nuclides, and cross-section identifiers had not been fixed for all materials used in the simulations, so participants made their own (different) choices, which caused some of the discrepancies initially noted.

3.4. Issues with the timetable of the exercises

Most exercises were planned with a timetable, which in almost all cases proved to be too ambitious and optimistic. This was partly because for many exercises the initial number of participants was lower than was expected and considered adequate for the purpose, so submission deadlines were postponed several times to increase participation after further publicity for the exercises. Further deadline extensions became necessary at the request of the participants.

After an initial analysis of the submitted solutions, participants whose results differed by more than expected from either the reference solution or from most other participants' solutions (as appropriate) were informed on this fact and invited to revise their solutions. Deadlines for the submission of revised results also had to be postponed several times.

As a result, the total duration of the exercises exceeded the typical length of stay of junior researchers at a given institute, making it difficult, if not impossible, to follow-up on abnormal results in some cases.

4. Lessons learnt

This section discusses the insights gained from the exercises from the perspective of the organizers.

4.1. Problem specification

The participants, as well as the organisers of the exercises, are committed to EURADOS and the intercomparisons in addition to their daily work. Therefore, the topics of the intercomparison exercises must be relevant to the participants' fields of work, and the tasks to be solved should not be overly demanding in terms of setup time. CPU resource requirements may also need to be considered but should generally be less of an issue.

To meet the workload requirements, some of the exercises presented in this Special Issue were designed with simplified idealistic geometric setups and irradiation conditions. Examples include the voxel phantom exercises for monoenergetic point sources, the uncertainty exercise in micro- and nanodosimetry (idealized energy spectrum), and the nanoparticle exercise (simplistic geometry). These simplifications have sometimes raised concerns among reviewers about the usefulness of the respective comparisons but seem justified given the aforementioned time constraints.

Regardless of the complexity of a task, a complete description of the problem to be solved with all relevant information must always be given. For Monte Carlo simulation exercises, this means that the radiation source, simulation geometry and materials must be comprehensively specified. On the other hand, it should generally not be specified exactly how the Monte Carlo simulation or the unfolding are to be carried out. The path to the solution, as well as the tools to be used, must be decided at the discretion of the participant. The participant must determine, for example, whether and which variance reduction techniques can or should be used, whether the transport of secondary charged particles should be simulated or how the thermal neutron transport should be performed.

However, depending on the aim of the exercise, a more detailed specification of intermediate steps or procedures may be advisable. For instance, whenever the performance of codes or their differences is to be assessed, it may be wise also to specify some of the aforementioned aspects of the simulations to ensure that the differences between results from different participants only reflect the differences in the codes that one is interested in.

4.2. Reporting of results

The task definition should include very precise instructions for reporting results, and templates should be provided where possible. Providing such a template, where the participants were requested to fill in their results in a pre-defined format, not only helps clarify exactly what output is required from them in each case, but also greatly facilitated the evaluation of the results in the respective exercises. This is especially important when spectral information is to be reported, where a lack of specification of bin division can make synopsis quite cumbersome when different participants use different bin sizes and/or linear and logarithmic equidistant bins.

In addition, asking for redundant information can help to identify potential problems with participants' data. Examples of this were: the voxel phantom exercise for internal dosimetry, where absorbed fractions and specific absorbed fractions were to be reported (differing by only one factor, i.e. organ mass); or the voxel phantom exercise for the X-ray examinations, where the results were to be reported normalised to both kerma free-in-air and kerma area product, which again differ by only one factor. In the case of the nanoparticle exercise, only reporting of results normalised to the number of primary particles was required. If normalisation to the area density of the emitted primary photons from the source had also been reported, the incorrect implementations of the simulation geometry would have been detected much earlier during the exercise.

4.3. Timing of the exercise

Regarding the problems encountered with non-responding participants at the revision of results stage, it is planned to set up rules in future exercises to get a more formal commitment from participants. The rules to be established concern deadlines, participation in the feedback loop,

and requirements for co-authorship to potential manuscripts (see Supplementary Fig. S1 and Supplementary Fig. S2).

In addition, more timely feedback to the participants might improve their preparedness to disclose details of their computational procedures and improvements. Long feedback intermissions make it difficult for the participants to recall exactly what was done and even what the exercise was about. It should be kept in mind that the organizers, as well as all participants, are performing these exercises alongside their daily duties.

In the reanalysis of the nanoparticle exercise, a set of hierarchical MS Excel templates were used that allowed a fast assessment of the internal consistency of the results reported by participants (in an Excel-template provided to them) as well as a 'live' synopsis via hyperlinks. As an illustration of this approach, Supplementary Fig. S3 to Supplementary Fig. S6 show screenshots of the "Synopsis" worksheets with easy-to-assess graphs and calculated figures of merit (integral quantities normalised so that their expected values are close to unity).

Using these templates to assess the consistency of a participant's results took only a few minutes and required only copying and pasting the results from the Excel templates completed by the participant into the template used for the analysis. In many cases, this enabled feedback in less than an hour. Of course, it was more time-consuming to identify the more sophisticated deviations from the exercise specifications.

Another issue with timing is that calculations with voxel phantoms may require large amounts of time. With some codes, even the visualisation of them can take up to several hours, and the production calculations even weeks. Some codes have ability to skip the geometry check at the beginning of production calculations (e.g. DBCN card in MCNP) and in this way speed up the calculations significantly.

4.4. Quality assurance of results

As indicated in Section 3.2, in all exercises some of the participants seemed to have submitted their results without first carrying out adequate quality control of their solutions, e.g. by simple plausibility checks or by comparison with literature data, if available. Approaches such as the one mentioned in Section 4.2 could allow for faster identification of outliers and more timely feedback to participants. This could alleviate some of the problems, such as where there was a lack of response from participants following feedback regarding abnormal results.

However, a significant degradation in the quality of results may persist, as the evolution of many computational tools towards greater ease of use also allows their use without a certain level of expertise, which may still be required for meaningful results. Interaction with some exercise participants who reported unreasonable results revealed a lack of understanding of several fundamental aspects. For example, that the results of calculations are not just numbers, but physical quantities (which have dimensions).

Some participants were also unaware that there are different ways of specifying the categorical variable of a histogram (lower or upper limit of the bin or the bin centre), which can vary between different codes and affect the comparison of results, such as when reported with different definitions of the meaning of the values on the x-axis. This could be countered by requiring both the lower and upper bin limits to be reported.

There were also cases where participants determined ratios with a finer bin size than specified in the task, and then re-binned their results for reporting by averaging the ratios over the larger bins instead of calculating the ratio between the sum of the numerators and the sum of the denominators. Detection of such elementary mistakes requires access to the original results, so their origins were not always immediately apparent to the organizers.

Considering that most exercises will lead to publications in the form of EURADOS reports or journal articles, compliance with the principles of FAIR data (an acronym for findability, accessibility, interoperability and reusability (Wilkinson et al., 2016),) is also a matter that should be

given more emphasis in the future. This is, of course, the responsibility of each participant, but appropriate commitments can be included in the application forms (Supplementary Fig. S1 and Supplementary Fig. S2). As a minimum requirement, participants should provide documentation on where the following files are stored and backed up:

– data files containing the reported results
– data files with the raw simulation output
– files used for their production (material data or other input files, code, etc.)
– log files and other supplementary output files of the simulations.

This information is essential when participants need to review their work for possible errors. However, in the exercises, delayed responses from participants were sometimes explained by difficulties in finding data, uncertainty about which version of a code was used, and similar such problems. Therefore, it may be useful for the organisers to collect this information - or even the files containing the metadata of the simulations - as part of the reporting of the results. This might be the case especially for participants who are inexperienced users of simulation codes, who may not be aware that the log files etc. generated by their codes contain information that complements their simulation results and is important for their quality.

It was not uncommon for files of original simulation results, which were shared by participants with the organisers during the feedback loops, to contain only columns of numbers. A header indicating which quantities are listed and which units have been used, and ideally also containing information on the code used and its version, as well as the date when the file was written, would be minimum requirements for the useability of these data files. In addition, the aforementioned auxiliary information is also needed.

5. Conclusions

Beyond doubt, the reported EURADOS exercises are beneficial to the field of computational dosimetry. They directly contribute to the training of the participants by improving their computational procedures through feedback with the task organisers. They lead also to the availability of representative dose values for various exposure conditions that may aid future novice users in the quality assurance of their methods. In addition, they also provide a snapshot of how well (or otherwise) the computational techniques are being applied within the community in general, and how well some of the concepts recommended by organizations such as ICRP are understood; the observed difficulties in correctly defining and evaluating effective doses (Eakins et al., 2021; Huet et al., 2022), or in determining bone marrow doses (Zankl et al., 2021a), are clear examples of the latter.

The obvious question of what could be done better in future exercises has been partly addressed in Section 4. A general answer to this question is not easy, since it depends on the objective of the exercises. The question will therefore continue to be the subject of discussion within EURADOS Working Group 6 when new exercises are prepared.

To avoid participants wasting their time on the tasks of an exercise in cases where they are prone to give incorrect results, due to a wrong idea of the task or ignorance of the dosimetric quantities to be determined, several modifications of the exercises can be considered. One could be to define explicitly the dosimetric quantities and normalisation quantities to be used. Another possibility would be to include in the definition of the task a list of checks to be made by the participants on the results, or even provide them with templates like those used in the reanalysis of the nanoparticle exercise.

Such changes to the exercises would mitigate the risk of potential errors by the participants. However, while this closer guidance may lead to improved results, this better agreement will only reflect the participants' capability to follow detailed instructions and not their actual state of expertise or their performance in real-world applications. Moreover,

concepts like effective dose are widely used and already well-defined elsewhere; arguably, it should not therefore be the role of EURADOS WG6, which focusses on computational dosimetry, to coach radiation professionals on the basic concepts that underpin radiological protection.

Most of the exercises were the types of task that the participants may be confronted with in their professional activities or in research. Generally, they then would have to perform the simulation or unfolding, and calculate the dosimetric quantities of interest, without specific guidance. A supervisor of an early-stage researcher, or a reviewer of potential papers arising from such work, can be a source of feedback that hopefully reveals major non-plausibilities in the results, if any. However, as discussed by (Rabus et al., 2021c), these potential quality filters often seem to have failed for simulation studies on nanoparticles so far.

A better option could therefore be to create a questionnaire to assess the knowledge of the participants, and then give more detailed instructions to the less experienced participants. One could also start the exercise with a webinar explaining what is expected, what participants should do, and explaining the importance of quality assurance. A recording of the webinar could also be made available on the EURADOS website so that participants who join the exercise later can refer to it.

Another potential improvement for some of the exercises, if repeated, could be to include several reporting steps in the exercise. For example, in the voxel phantom exercise for the X-ray examinations, participants could first need to report their results for the position of the radiation source in relation to the phantom; they could then get feedback if it is outside the uncertainty band of the reference value and be invited to report a revised value. On request they could also get the correct positions as feedback and run their simulations for these. Alternatively, such future exercises could have a first step with as few specifications as possible, a second with plausibility checks to be performed by the participants, and a third with feedback on the submitted results and a request for revision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.radmeas.2022.106822.

References

Broggio, D., Bento, J., Caldeira, M., Cardenas-Mendez, E., Farah, J., Fonseca, T., Konvalinka, C., Liu, L., Perez, B., Capello, K., Cowan, P., Cruzate, J.-A., Freire, L., Gómez-Ros, J.-M., Gossio, S., Heide, B., Huikari, J., Hunt, J., Kinase, S., Kramer, G. H., Kurihara, O., Kyrieleis, A., Lebacq, A.-L., Leone, D., Li, C., Li, J., Mihailescu, L.-C., Moraleda, M., Navarro, J.-F., Oliveira, C., Puerta, N., Reichelt, U., Simões, C., Sommer, D., Takahashi, M., Teles, P., Vanhavere, F., Vrba, T., Franck, D., Gualdrini, G., Lopez, M.-A., 2012. Monte Carlo modelling for the in vivo lung monitoring of enriched uranium: results of an international comparison. Radiat. Meas. 47, 492–500. https://doi.org/10.1016/j.radmeas.2012.04.020.

Caccia, B., Le Roy, M., Blideanu, V., Andenna, C., Arun, C., Czarnecki, D., El Bardouni, T., Gschwind, R., Huot, N., Martin, E., Zink, K., Zoubair, M., Price, R., De Carlan, L., 2017. EURADOS intercomparison exercise on Monte Carlo modelling of a medical

linear accelerator. Ann. Dell Ist. Super. Sanita 53, 314–321. https://doi.org/10.4415/ANN_17_04_07.

De Saint-Hubert, M., Farah, J., Klodowska, M., Romero-Expósito, M.T., Tyminska, K., Mares, V., Olko, P., Stolarczyk, L., Trinkl, S., 2022. The influence of nuclear models and Monte Carlo radiation transport codes on stray neutron dose estimations in proton therapy. Radiat. Meas. 150, 106693 https://doi.org/10.1016/j.radmeas.2021.106693.

De Saint-Hubert, M., Tymińska, K., Stolarczyk, L., Brkić, H., 2021. Fetus dose calculation during proton therapy of pregnant phantoms using MCNPX and MCNP6.2 codes. Radiat. Meas. 149, 106665 https://doi.org/10.1016/j.radmeas.2021.106665.

Eakins, J., Huet, C., Brkić, H., Capello, K., Desorgher, L., Epstein, L., Hunt, J.G., Kim, H. S., Krstic, D., Lee, Y.-K., Manohari, M., Nikezic, D., Shukrun, R.H., Souza-Santos, D., Tymińska, K., 2021. Monte Carlo calculation of organ and effective dose rates from ground contaminated by Am-241: results of an international intercomparison exercise. Radiat. Meas. 148, 106649 https://doi.org/10.1016/j.radmeas.2021.106649.

Gómez-Ros, J.M., Bedogni, R., Domingo, C., Eakins, J.S., Roberts, N., Tanner, R.J., 2022. Results of the EURADOS international comparison exercise on neutron spectra unfolding in Bonner spheres spectrometry. Radiat. Meas. 153, 106755 https://doi.org/10.1016/j.radmeas.2022.106755.

Gómez-Ros, J.M., Bedogni, R., Domingo, C., Eakins, J.S., Roberts, N., Tanner, R.J., 2018. International comparison exercise on neutron spectra unfolding in bonner spheres spectrometry: problem description and preliminary analysis. Radiat. Protect. Dosim. 180, 70–74. https://doi.org/10.1093/rpd/ncy002.

Gómez-Ros, J.M., Moraleda, M., Arce, P., Bui, D.-K., Dang, T.-M.-L., Desorgher, L., Kim, H.S., Krstic, D., Kuć, M., Le, N.-T., Lee, Y.-K., Nguyen, N.-Q., Nikezic, D., Tymińska, K., Vrba, T., 2021. Monte Carlo calculation of the organ equivalent dose and effective dose due to immersion in a 16N beta source in air using the ICRP reference phantoms. Radiat. Meas. 145, 106612 https://doi.org/10.1016/j.radmeas.2021.106612.

Gualdrini, G., Agosteo, S., Ménard, S., Price, R.A., Chartier, J.-L., Großwendt, B., Kodeli, I., Leuthold, G.P., Siebert, B.R.L., Tagziria, H., Tanner, R.J., Terrissol, M., Zankl, M., 2005. QUADOS intercomparison: a summary of photon and charged particle problems. Radiat. Protect. Dosim. 115, 587–599. https://doi.org/10.1093/rpd/nci197.

Gualdrini, G., Tanner, R.J., Agosteo, S., Pola, A., Bedogni, R., Ferrari, P., Lacoste, V., Bordy, J.-M., Chartier, J.-L., de Carlan, L., Gomez Ros, J.-M., Großwendt, B., Kodeli, I., Price, R.A., Rollet, S., Schultz, F., Siebert, B., Terrissol, M., Zankl, M., 2008. Analysis of the CONRAD computational problems expressing only stochastic uncertainties: neutrons and protons. Radiat. Protect. Dosim. 131, 7–14. https://doi.org/10.1093/rpd/ncn241.

Harrison, R.M., Ainsbury, E., Alves, J., Bottollier-Depois, J.-F., Breustedt, B., Caresana, M., Clairand, I., Fantuzzi, E., Fattibene, P., Gilvin, P., Hupe, O., Knežević, Ž., Lopez, M.A., Olko, P., Olšovcová, V., Rabus, H., Rühm, W., Silari, M., Stolarczyk, L., Tanner, R., Vanhavere, F., Vargas, A., Woda, C., 2021. Eurados strategic research agenda 2020: vision for the dosimetry of ionising radiation. Radiat. Protect. Dosim. 194, 42–56. https://doi.org/10.1093/rpd/ncab063.

Huet, C., Eakins, J., Zankl, M., Gómez-Ros, J.M., Jansen, J., Moraleda, M., Struelens, L., Akar, D.K., Borbinha, J., Brkić, H., Bui, D.K., Capello, K., Linh Dang, T.M., Desorgher, L., Di Maria, S., Epstein, L., Faj, D., Fantinova, K., Ferrari, P., Gossio, S., Hunt, J., Jovanovic, Z., Kim, H.S., Krstic, D., Le, N.T., Lee, Y.-K., Murugan, M., Nadar, M.Y., Nguyen, N.-Q., Nikezic, D., Patni, H.K., Santos, D.S., Tremblay, M., Trivino, S., Tymińska, K., 2022. Monte Carlo calculation of organ and effective doses due to photon and neutron point sources and typical X-ray examinations: results of an international intercomparison exercise. Radiat. Meas. 150, 106695 https://doi.org/10.1016/j.radmeas.2021.106695.

ICRP, 2009. ICRP publication 110: adult reference computational phantoms. Ann. ICRP 39, 1—164. https://doi.org/10.1016/j.icrp.2009.09.001.

ICRP, N., 2010. ICRP Publication 116. Conversion coefficients for radiological protection quantities for external radiation exposures. Ann. ICRP 40, 1–257. https://doi.org/10.1016/j.icrp.2011.10.001.

Li, W.B., Belchior, A., Beuve, M., Chen, Y.Z., Maria, S.D., Friedland, W., Gervais, B., Heide, B., Hocine, N., Ipatov, A., Klapproth, A.P., Li, C.Y., Li, J.L., Multhoff, G., Poignant, F., Qiu, R., Rabus, H., Rudek, B., Schuemann, J., Stangl, S., Testa, E., Villagrasa, C., Xie, W.Z., Zhang, Y.B., 2020a. Intercomparison of dose enhancement ratio and secondary electron spectra for gold nanoparticles irradiated by X-rays calculated using multiple Monte Carlo simulation codes. Phys. Med. 69, 147–163. https://doi.org/10.1016/j.ejmp.2019.12.011.

Li, W.B., Beuve, M., Maria, S.D., Friedland, W., Heide, B., Klapproth, A.P., Li, C.Y., Poignant, F., Rabus, H., Rudek, B., Schuemann, J., Villagrasa, C., 2020b. Corrigendum to "Intercomparison of dose enhancement ratio and secondary electron spectra for gold nanoparticles irradiated by X-rays calculated using multiple Monte Carlo simulation codes" [Phys. Med. 69. Phys. Med. 80, 383–388. https://doi.org/10.1016/j.ejmp.2020.10.008 (2020) 147-163].

Price, R.A., Gualdrini, G., Agosteo, S., Ménard, S., Chartier, J.-L., Großwendt, B., Kodeli, I., Leuthold, G.P., Siebert, B.R.L., Tagziria, H., Tanner, R.J., Terrissol, M., Zankl, M., 2006. Pitfalls and modelling inconsistencies in computational radiation dosimetry: lessons learnt from the QUADOS intercomparison. Part II: photons, electrons and protons. Radiat. Protect. Dosim. 118, 155–166. https://doi.org/10.1093/rpd/ncl014.

Rabus, H., Gómez-Ros, J.M., Villagrasa, C., Eakins, J., Vrba, T., Blideanu, V., Zankl, M., Tanner, R., Struelens, L., Brkić, H., Domingo, C., Baiocco, G., Caccia, B., Huet, C., Ferrari, P., 2021a. Quality assurance for the use of computational methods in dosimetry: activities of EURADOS Working Group 6 'Computational Dosimetry. J. Radiol. Prot. 41, 46–58. https://doi.org/10.1088/1361-6498/abd914.

Rabus, H., Li, W.B., Nettelbeck, H., Schuemann, J., Villagrasa, C., Beuve, M., Di Maria, S., Heide, B., Klapproth, A.P., Poignant, F., Qiu, R., Rudek, B., 2021b. Consistency checks of results from a Monte Carlo code intercomparison for emitted electron spectra and energy deposition around a single gold nanoparticle irradiated by X-rays. Radiat. Meas. 147, 106637 https://doi.org/10.1016/j.radmeas.2021.106637.

Rabus, H., Li, W.B., Villagrasa, C., Schuemann, J., Hepperle, P.A., de la Rosales, L.F., Beuve, M., Maria, S.D., Klapproth, A.P., Li, C.Y., Poignant, F., Rudek, B., Nettelbeck, H., 2021c. Intercomparison of Monte Carlo calculated dose enhancement ratios for gold nanoparticles irradiated by X-rays: assessing the uncertainty and correct methodology for extended beams. Phys. Med. 84, 241–253. https://doi.org/10.1016/j.ejmp.2021.03.005.

Rühm, W., Ainsbury, E., Breustedt, B., Caresana, M., Gilvin, P., Knežević, Ž., Rabus, H., Stolarczyk, L., Vargas, A., Bottollier-Depois, J.F., Harrison, R., Lopez, M.A., Stadtmann, H., Tanner, R., Vanhavere, F., Woda, C., Clairand, I., Fantuzzi, E., Fattibene, P., Hupe, O., Olko, P., Olšovcová, V., Schuhmacher, H., Alves, J.G., Miljanic, S., 2020. The European radiation dosimetry group – review of recent scientific achievements. Radiat. Phys. Chem. 168, 108514. https://doi.org/10.1016/j.radphyschem.2019.108514.

Rühm, W., Bottollier-Depois, J.F., Gilvin, P., Harrison, R., Knežević, Ž., Lopez, M.A., Tanner, R., Vargas, A., Woda, C., 2018. The work programme of EURADOS on internal and external dosimetry. Ann. ICRP 47, 20–34. https://doi.org/10.1177/0146645318756224.

Siebert, B.R.L., Tanner, R.J., Chartier, J.-L., Agosteo, S., Großwendt, B., Gualdrini, G., Ménard, S., Kodeli, I., Leuthold, G.P., Price, R.A., Tagziria, H., Terrissol, M., Zankl, M., 2006. Pitfalls and modelling inconsistencies in computational radiation dosimetry: lessons learnt from the QUADOS intercomparison. Part I: neutrons and uncertainties. Radiat. Protect. Dosim. 118, 144–154. https://doi.org/10.1093/rpd/ncl013.

Tanner, R.J., Chartier, J.-L., Siebert, B.R.L., Agosteo, S., Großwendt, B., Gualdrini, G., Kodeli, I., Leuthold, G.P., Ménard, S., Price, R.A., Tagziria, H., Terrissol, M., Zankl, M., 2004. Intercomparison on the usage of computational codes in radiation dosimetry. Radiat. Protect. Dosim. 110, 769–780. https://doi.org/10.1093/rpd/nch228.

Villagrasa, C., Bordage, M.-C., Bueno, M., Bug, M., Chiriotti, S., Gargioni, E., Heide, B., Nettelbeck, H., Parisi, A., Rabus, H., 2019. Assessing the contribution of cross-sections to the uncertainty of Monte Carlo calculations in micro- and nanodosimetry. Radiat. Protect. Dosim. 183, 11–16. https://doi.org/10.1093/rpd/ncy240.

Villagrasa, C., Rabus, H., Baiocco, G., Perrot, Y., Parisi, A., Struelens, L., Qiu, R., Beuve, M., Poignant, F., Pietrzak, M., Nettelbeck, H., 2022. Intercomparison of micro- and nanodosimetry Monte Carlo simulations: an approach to assess the influence of different cross-sections for low-energy electrons on the dispersion of

results. Radiat. Meas. 150, 106675 https://doi.org/10.1016/j.radmeas.2021.106675.

Vrba, T., Broggio, D., Caldeira, M., Capello, K., Fantínová, K., Franck, D., Gómez-Ros, J. M., Hunt, J., Kinase, S., Leone, D., Lombardo, P.A., Manohari, M., Marzocchi, O., Moraleda, M., Nogueira, P., Ośko, J., Arron, S., Suhl, S., Takahashi, M., Teles, P., Tremblay, M., Tymińska, K., Lopez, M.A., Tanner, R., 2015. EURADOS intercomparison exercise on MC modelling for the in-vivo monitoring of AM-241 in skull phantoms (Part II and III). Radiat. Phys. Chem. 113, 59–71. https://doi.org/10.1016/j.radphyschem.2015.04.009.

Vrba, T., Nogueira, P., Broggio, D., Caldeira, M., Capello, K., Fantínová, K., Figueira, C., Hunt, J., Leone, D., Murugan, M., Marzocchi, O., Moraleda, M., Shutt, A., Suh, S., Takahashi, M., Tymińska, K., Lopez, M.A., Tanner, R., 2014. EURADOS intercomparison exercise on MC modeling for the in-vivo monitoring of Am-241 in skull phantoms (Part I). Radiat. Phys. Chem. 104, 332–338. https://doi.org/10.1016/j.radphyschem.2013.12.010.

Wilkinson, M.D., Dumontier, M., Aalbersberg, IjJ., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3, 160018. https://doi.org/10.1038/sdata.2016.18.

Zankl, M., Eakins, J., Gómez Ros, J.-M., Huet, C., 2021a. The ICRP recommended methods of red bone marrow dosimetry. Radiat. Meas. 146, 106611 https://doi.org/10.1016/j.radmeas.2021.106611.

Zankl, M., Eakins, J., Gómez Ros, J.-M., Huet, C., Jansen, J., Moraleda, M., Reichelt, U., Struelens, L., Vrba, T., 2021b. EURADOS intercomparison on the usage of the ICRP/ICRU adult reference computational phantoms. Radiat. Meas. 145, 106596 https://doi.org/10.1016/j.radmeas.2021.106596.

Zankl, M., Gómez Ros, J.-M., Moraleda, M., Reichelt, U., Akar, D.K., Borbinha, J., Desorgher, L., Di Maria, S., El Bakkali, J., Fantinova, K., Ferrari, P., Gossio, S., Hunt, J., Jovanovic, Z., Kim, H.S., Krstic, D., Lee, Y.-K., Nadar, M.Y., Nikezic, D., Patni, H.K., Murugan, M., Triviño, S., 2021c. Monte Carlo calculation of organ dose coefficients for internal dosimetry: results of an international intercomparison exercise. Radiat. Meas. 148, 106661 https://doi.org/10.1016/j.radmeas.2021.106661.