A quantitative assessment of Geant4 for predicting the yield and distribution of positron-emitting fragments in ion beam therapy

Andrew Chacon[1], Harley Rutherford[1,2], Akram Hamato[3], Munetaka Nitta[3], Fumihiko Nishikido[3], Yuma Iwao[3], Hideaki Tashima[3], Eiji Yoshida[3], Go Akamatsu[3], Sodai Takyu[3], Han Gyu Kang[3], Daniel R. Franklin[4], Katia Parodi[5], Taiga Yamaya[3], Anatoly Rosenfeld[2,6], Susanna Guatelli[2,6], Mitra Safavi-Naeini[1,2,7]

[1] Australian Nuclear Science and Technology Organisation (ANSTO), NSW, Australia
[2] Centre for Medical Radiation Physics, University of Wollongong, Wollongong NSW 2522 Australia
[3] National Institutes for Quantum Science and Technology, Chiba, Japan
[4] School of Electrical and Data Engineering, University of Technology Sydney, Australia
[5] Ludwig-Maximilians-Universität München, Department of Medical Physics, Faculty of Physics, Garching b. Munich, Germany
[6] Illawarra Health and Medical Research Institute, University of Wollongong, Wollongong NSW 2522 Australia
[7] Brain and Mind Centre, University of Sydney, Sydney, NSW, Australia

E-mail: `mitras@ansto.gov.au`

Abstract.

Objective: To compare the accuracy with which different hadronic inelastic physics models across ten Geant4 Monte Carlo simulation toolkit versions can predict positron-emitting fragments produced along the beam path during carbon and oxygen ion therapy.

Approach: Phantoms of polyethylene, gelatin, or poly(methyl methacrylate) were irradiated with monoenergetic carbon and oxygen ion beams. Post-irradiation, 4D PET images were acquired and parent ^{11}C, ^{10}C and ^{15}O radionuclides contributions in each voxel were determined from the extracted time activity curves. Next, the experimental configurations were simulated in Geant4 Monte Carlo versions 10.0 to 11.1, with three different fragmentation models - binary ion cascade (BIC), quantum molecular dynamics (QMD) and the Liege intranuclear cascade (INCL++) - 30 model-version combinations. Total positron annihilation and parent isotope production yields predicted by each simulation were compared between simulations and experiments using normalised mean squared error and Pearson cross-correlation coefficient. Finally, we compared the depth of the maximum positron annihilation yield and the distal point at which the positron yield decreases to 50% of peak between each model and the experimental results.

Main results: Performance varied considerably across versions and models, with no one version/model combination providing the best prediction of all positron-emitting fragments in all evaluated target materials and irradiation conditions. BIC in Geant4 10.2 provided the best overall agreement with experimental results in the largest number of test cases. QMD consistently provided the best estimates of both the depth of peak positron yield (10.4 and 10.6) and the distal 50%-of-peak point (10.2), while BIC also performed well and INCL generally performed the worst across most Geant4 versions.

Significance: The best predictions of the spatial distribution of positron annihilations and positron-emitting fragment production along the beam path during carbon and oxygen ion therapy was obtained using Geant4 10.2.p03 with BIC or QMD. These version/model combinations are recommended for future heavy ion therapy research.

1. Introduction

One of the chief advantages of particle therapy as a treatment for cancer is the high dose gradient between the treatment area and surrounding regions [1]. This precision necessitates the use of sophisticated treatment planning and quality assurance methods to ensure proper delivery of the prescribed dose to the target only. These methods, in turn, are heavily reliant on Monte Carlo simulation methods, which are used for modelling the interaction of high-energy charged particles with the patient.

Good models for nuclear fragmentation processes are especially critical for faithfully simulating imaging applications in particle therapy, such as PET-based dose estimation methods for quality assurance, since the production and distribution of positron-emitting radionuclide fragments directly affects the quality of the resulting image [2, 3, 4, 5]. One of the leading fully open source Monte Carlo toolkits for modelling the interaction of radiation and matter, **Geant4**, currently offers a choice of three hadronic inelastic fragmentation models that are appropriate for particle therapy - binary ion cascade (BIC), quantum molecular dynamics (QMD), and Liège intranuclear cascade (INCL++) [6, 7, 8]‡. In a previous study, we evaluated these models by comparing the spatial distributions of positron-emitting radionuclides predicted following irradiation of PMMA, gelatin and polyethylene targets by monoenergetic carbon and oxygen ion beams (simulated using Geant4 10.2.p03) to equivalent results estimated from experimentally-obtained PET data [10]. The BIC model was found to provide the best estimates overall; however, none of the models provided a perfect fit in all evaluated cases, and some significant discrepancies were observed.

Since the publication of our previous study, there have been several updates to Geant4; specifically, six minor releases (versions 10.x) and one major release (version 11, which has since been updated to version 11.1). Each of these releases includes modifications to the physics models implemented in Geant4, which can affect the simulation of positron-emitting fragment production in particle therapy.

In this work, we have extended our earlier study, and present a quantitative evaluation of Geant4's ability to predict positron-emitting fragment production across a total of ten different stable versions (10.0.p04, 10.1.p03, 10.2.p03, 10.3.p03, 10.4.p03, 10.5.p01, 10.6.p03, 10.7.p02, 11.0 and 11.1) which have followed the previous major release (10.0) for each of the three different fragmentation models [10]. In addition to the normalised mean squared error (NMSE) metric used in the previous study, three additional metrics - the Pearson cross-correlation coefficient (CCC), the depth of the positron annihilation peak, and the depth at which the positron annihilation intensity has decreased to 50% of the peak - are also used to compare the shape of the predicted positron-emitting fragment distributions with the experimentally measured distributions.

2. Materials and Methods

This section describes the methods used for obtaining and quantitatively comparing the experimental and simulated positron annihilation profiles. The general approach is similar to that used in our previous study (see [10]); however, it has been extended to include a much wider range of Geant4 versions, and additional comparison metrics are introduced.

The experimental methods used to estimate the total positron annihilation profile and activity of the dominant positron-emitting fragment isotopes (^{11}C, ^{10}C and ^{15}O) are briefly summarised in Section 2.1. Equivalent simulation configurations were constructed for each Geant4 version under test, and the total positron annihilation profile and activity of ^{11}C, ^{10}C and ^{15}O were predicted for each beam ion/energy, target material, hadronic inelastic fragmentation model and Geant4 version; the design and parameters of these simulations are described in detail in Section 2.2.

‡ INCL++ is considered the most appropriate option for neutron spallation simulations, but is included here for completeness [9].

Table 1. Beam parameters for each ion species and energy. The energy spread is 0.2 % of nominal energy in each case; 95% confidence intervals are given for beam flux.

Ion	**Energy (MeV/u)**	σ_x **(mm)**	σ_y **(mm)**	**Beam flux (pps)**
^{12}C	148.5	2.77	2.67	$1.8\times10^9 \pm 3.8\times10^7$
^{12}C	290.5	3.08	4.70	$1.8\times10^9 \pm 6.4\times10^7$
^{12}C	350	2.50	2.98	$1.8\times10^9 \pm 4.6\times10^7$
^{16}O	148	2.79	2.89	$1.1\times10^9 \pm 2.8\times10^7$
^{16}O	290	2.60	4.90	$1.1\times10^9 \pm 7.0\times10^7$

Results in each of the three target materials and 5 ion/energy combinations were then compared to those predicted in equivalent simulations performed in Geant4 using each of hadronic fragmentation models (BIC, QMD and INCL++) across the ten evaluated Geant4 versions for a total of 150 unique target/ion/energy/version/model test conditions. The total positron yields and yields of the individual positron-emitting fragment species from each model and Geant4 version were then compared with the experimental annihilation profiles using the following metrics in each of the entrance, build-up, and Bragg peak and tail regions:

- Normalised mean squared error (NMSE); and
- Pearson cross-correlation coefficient (CCC)

Additionally, the depth of the positron annihilation peak and the depth of the distal point at which the magnitude of the positron annihilation profile decreases to 50% of the peak value are evaluated. All metrics are described in detail in Section 2.3.

2.1. Experimental configuration

The experimental data obtained in our 2019 paper were used as the ground truth for this simulation study; a detailed description of the experimental procedures is presented in that paper [10]. In summary, phantoms constructed from either pure PMMA, polyethylene or gelatin (encased in a thin-walled PMMA container), each with dimensions of 100 mm×100 mm×300 mm, were irradiated with monoenergetic carbon or oxygen ion beams of various energies - three for carbon ions and two for oxygen (see Table 1). Positron annihilation profiles (with respect to depth in the target) were estimated across the full width at tenth maximum (FWTM) of the beam using the whole-body DOI-PET scanner prototype developed at QST [11]. These profiles were decomposed into the individual population of each of the dominant parent positron-emitting fragments (^{11}C, ^{10}C and ^{15}O) at $t = 0$ (end of irradiation period) by fitting the observed time-decay curves in each voxel to a multiexponential decay model.

2.2. Simulation parameters

The same beam parameters, phantom compositions and geometries used in the experimental measurements were modelled in each version of Geant4. Apart from minor modifications to the simulation source code required due to version-to-version changes in certain Geant4 application programming interfaces (APIs), the code was identical across versions. Simulations were performed using each of the 10 most recent stable releases of Geant4: 10.0.p04, 10.1.p03, 10.2.p03, 10.3.p03, 10.4.p03, 10.5.p01, 10.6.p03, 10.7.p02, 11.0 and 11.1. For brevity, the patch number will be dropped when referring to the version of Geant4.

For each version of Geant4, three alternative hadronic ion fragmentation models were evaluated - Binary Ion Cascade (BIC), Quantum Molecular Dynamics (QMD) and Liège Intranuclear Cascade (INCL) models§ [7, 8]. All simulations modelled electromagnetic interactions using the standard option 3 list (`G4EmStandardPhysics_option3`). The remaining physics processes, including hadronic physics models, are listed in Table 10.

The location of each positron annihilation, as well as the identity of the parent isotope which decayed to emit each positron (principally ^{11}C, ^{10}C and ^{15}O), were scored with a resolution of 1.5 mm^3 to match the voxel dimensions of the experimental OpenPET image reconstruction output. The pristine positron annihilation profiles were convolved with a 2.3 mm FWHM Gaussian filter to simulate the measured point spread function of the PET system [11].

A total of 20 runs, each with 10^8 primary particles were simulated for each version/model combination. In our previous work, we established that this is sufficient to limit the run-to-run ratio of standard deviation to mean across the build-up and Bragg peak region of the profiles to less than 5% ([10]). Each of the simulated profiles is randomly paired with one of the experimental profiles (for the same target, ion species and beam energy) and then the performance metrics are calculated, with the statistical distribution of each metric used to generate the confidence intervals shown in the results presented in the Supplementary Materials.

2.3. Evaluation methods and metrics

The irradiated target was divided into three separate regions for analysis since different physics processes dominate in each: the entrance region, the build-up and Bragg peak region, and the tail region. This segmentation is defined in the same way as in our previous paper [10]; in summary, the central build-up and Bragg peak region is defined as follows:

- The proximal edge in the z dimension (along the path of the beam) is defined as the first point at which the dose deposited along the central axis exceeds the entrance

---

§ The INCL model was developed specifically for spallation reactions but is included in this study as it can also model fragmentation.

plateau dose by more than 5% of the difference between peak dose and the entrance plateau dose; and

- The distal edge in z is defined as the last point at which the deposited dose is greater than 5% of the absolute peak dose value.

The entrance region is then defined as the region proximal to the build-up and Bragg peak region, while the tail region is defined as the region distal to the build-up and Bragg peak region.

The yields of the positron-emitting nuclei are defined by (1):

$$\text{Yield (Isotope)} = \frac{N\ (\text{Isotope})}{N\ (\text{Primary})} \tag{1}$$

where N (Isotope) is the yield of the isotope under study in that region and N (Primary) is the total number of primary particles. Yields were calculated for each voxel along the beam path.

Three different metrics were chosen to quantify the accuracy of each model in Geant4: the normalised mean squared error (NMSE), the Pearson cross-correlation coefficient (CCC), and the range (depth along the path of the beam) of both the positron annihilation peak and the point beyond the peak at which positron annihilation intensity decreases by 50%.

NMSE measures the average squared difference between the experimental measurements and simulation-predicted positron yields in each region. NMSE is most useful in regions of relatively high yield (especially in the entrance and build-up and Bragg peak regions); the relatively low statistics available in the tail region limit the value of the NMSE there.

NMSE is defined as:

$$NMSE = \frac{\displaystyle\sum_{i=1}^{N_{reg}} |S_i - E_i|^2}{\displaystyle\sum_{i=1}^{N_{reg}} |E_i|^2} \tag{2}$$

where S_i and E_i are the simulation and experimental yields in the ith voxel of the N_{reg} voxels in region reg (with a lower value indicating a better match).

For the NMSE metric, we identify the best-performing model (with the lowest mean NMSE) and consider any other model whose mean NMSE is within two standard deviations of the best-performing version/model as being statistically equal. For a Gaussian random distribution, this would correspond to a 95% confidence interval (although, as can be seen in the box plots of the NMSE results included in the Supplementary Materials, the NMSE distributions often deviate from the Gaussian model).

The Pearson cross-correlation coefficient compares the degree of linear dependence of one profile to another - that is, the degree to which changes in the profiles occur at the same location and in the same direction. Thus, the Pearson CCC quantifies the differences in shape between the simulation-predicted positron-emitting fragment distributions and the experimental measurements, without regard to differences in the magnitude of the profiles. The Pearson CCC is defined as:

$$CCC = \frac{\sum_{i=1}^{N_{reg}} (S_{norm,i} - \overline{S_{norm}})(E_i - \overline{E_{norm}})}{\sqrt{(\sum_{i=1}^{N_{reg}} (S_{norm,i} - \overline{S_{norm}})^2)(\sum_{i=1}^{N_{reg}} (E_{norm,i} - \overline{E_{norm}})^2)}} \tag{3}$$

where $S_{norm,i}$ and $E_{norm,i}$ are the normalised simulation and experimental yields in the ith voxel of the N_{reg} voxels in region reg. Normalisation is performed by dividing each S_i and E_i by the maximum value in its respective region. $\overline{S_{norm}}$ and $\overline{E_{norm}}$ are the mean values in each region.

When comparing the models, the closer that the CCC between the simulation output and the experimental estimate of positron-emitting fragment distribution is to +1, the more accurate the prediction. A Pearson CCC greater than +0.8 is generally considered to be "very strong" [12]. In this work, we aim to identify the very best version/model combinations; therefore, a Pearson CCC threshold of 0.95 is chosen to identify those combinations which have produced exceptionally good predictions of the shape of the yield profiles. It is important to note that this threshold is quite arbitrary, and the most appropriate threshold depends on the application; readers are referred to the Supplementary Data for the complete set of results.

For each version of Geant4, phantom, beam type and energy, the NMSE and CCC were calculated for both total annihilation photon yield profiles and also for the profiles of the three main positron-emitting fragment species (^{10}C, ^{11}C and ^{15}O). The calculation was repeated for each of the N_{reg} regions (entrance, build-up and Bragg peak, and tail regions). The NMSE and the CCC were then compared across all evaluated Geant4 versions for each region, phantom material and beam type.

A total of 5 energy/ion combinations are evaluated (carbon ions at three energies and oxygen ions at two energies). For oxygen ions, three target materials (gelatine, PMMA and polyethylene) are evaluated for total positron annihilation yield and ^{11}C/^{10}C/^{15}O yield. For carbon ions, the same three target materials are evaluated for total positron annihilation yield and ^{11}C/^{10}C yield and two for ^{15}O yield (polyethylene is omitted since it is not possible to produce ^{15}O fragments with a ^{12}C ion beam and a PE target which only contains carbon and hydrogen). Thus, a total of 15 cases are evaluated for total positron annihilation, ^{11}C yield and ^{10}C yield, while 12 are cases evaluated for ^{15}O yield.

For range calculations, the difference between the depths at which the positron annihilation yield reached its maximum value in the experiment and simulation was calculated (see (4)). Additionally, the point distal to this maximum at which positron annihilation yield decreases to 50% of the maximum value was also compared between experiment and simulation. For each version and model, the mean differences between the experimental and simulation-based values, as well as the standard deviations and

Table 2. Number of test cases for which each Geant4 version/model combination achieved the lowest or equal-lowest NMSE in the **entrance region**. **Bold text** denotes the version/model achieving the highest (or equal-highest) number of best results for each combination of ion/energy/target.

Version	**Total**			**^{11}C**			**^{10}C**			**^{15}O**		
	BIC	**QMD**	**INCL**	**BIC**	**QMD**	**INCL**	**BIC**	**QMD**	**INCL**	**BIC**	**QMD**	**INCL**
10	**6**	0	0	**11**	3	2	0	0	2	6	2	0
10.1	**6**	0	0	**11**	3	2	0	0	3	6	2	0
10.2	5	0	0	**11**	3	2	0	0	3	5	2	0
10.3	**6**	0	0	6	0	0	5	3	6	4	2	0
10.4	**6**	0	0	1	0	0	2	3	6	**9**	2	0
10.5	2	0	0	0	0	0	2	5	**8**	3	1	0
10.6	4	0	0	1	0	0	0	0	5	5	4	0
10.7	3	0	0	1	0	0	0	0	5	5	2	0
11	4	0	0	1	0	0	0	0	5	5	2	0
11.1	4	0	0	0	0	0	0	0	5	5	2	0

maximum differences were calculated across all test cases (ion species, energies and target materials).

$$\delta_{voxel} = R_{simulation} - R_{experiment} \tag{4}$$

where R_x is the range (depth) of the voxel with the maximum value (or, for distal 50%-of-peak, the first distal voxel to fall below 50% of the maximum value) in either the simulation or experiment.

3. Results and Discussion

The number of cases in which each version/model combination performed the best or equal-best in terms of each of the evaluated metrics are counted across all simulations in the entrance, build-up and Bragg peak and tail regions, and summarised in this section. Detailed results for each experiment are included in the Supplementary Materials.

3.1. Entrance region

In the entrance region, positron-emitting fragments are created by target fragmentation rather than projectile fragmentation. The projectile ions lose energy via Coulomb interactions, slowing down at an approximately constant rate as they traverse this region, with only gradual changes to projectile/target cross sections. As a result, the positron-emitting fragment distributions are expected to exhibit an approximately flat depthwise profile in this region.

NMSE and Pearson CCC results between simulation and experimental total positron annihilation profiles in the entrance region are summarised in Tables 2 and 3, respectively, with corresponding figures shown in Supplementary Material Section 1.

Table 3. Number of test cases for which each Geant4 version/model combination achieved a CCC greater than 0.95 in the **entrance region**. **Bold text** denotes the version/model achieving the highest number of best results for each combination of ion/energy/target.

Version	**Total**			**^{11}C**			**^{10}C**			**^{15}O**		
	BIC	**QMD**	**INCL**	**BIC**	**QMD**	**INCL**	**BIC**	**QMD**	**INCL**	**BIC**	**QMD**	**INCL**
10	2	0	1	**4**	0	0	**2**	**2**	0	**3**	0	**3**
10.1	1	0	1	**4**	0	0	**2**	**2**	0	**3**	0	**3**
10.2	2	0	1	3	0	0	**2**	**2**	0	**3**	0	**3**
10.3	2	0	**3**	3	1	3	**2**	**2**	0	**3**	0	2
10.4	2	0	**3**	**4**	1	3	1	**2**	0	**3**	0	2
10.5	**3**	0	1	3	2	**4**	1	**2**	0	**3**	0	**3**
10.6	**3**	0	1	2	1	2	1	**2**	0	**3**	0	2
10.7	**3**	0	1	3	1	3	1	**2**	0	**3**	0	2
11	**3**	0	1	2	1	3	1	**2**	0	**3**	0	2
11.1	1	0	1	2	1	2	1	**2**	0	**3**	0	2

For the entrance region, the BIC model implemented in Geant4 versions 10, 10.1, 10.3 and 10.4 provided the (equal) lowest NMSE of the yields of total positron annihilation in 5 out of 15 cases. The BIC model in Geant4 10, 10.1 and 10.2 also provided the (equal) lowest NMSE for ^{11}C fragment production (11/15 cases), whereas for ^{10}C the best version/model combination was 10.5/INCL (8/15 cases) and for ^{15}O it was 10.6/BIC (9/12 cases).

Geant4 versions 10.5-11 with BIC and 10.3/10.4 with INCL each achieved a Pearson CCC greater than 0.95 (3/15 cases) for total positron yield; QMD did not reach the threshold for any test case in any version of Geant4.

Results for individual radionuclides were also mixed, with 10/BIC, 10.1/BIC, 10.4/BIC and 10.5/INCL achieving the threshold in 4/15 cases for ^{11}C, 10-10.4/BIC and all versions with QMD reaching the threshold in 2/15 cases for ^{10}C, and all versions with BIC and 10/INCL, 10.1/INCL, 10.2/INCL, and 10.5/INCL reaching the threshold for ^{15}O.

3.2. Build-up and Bragg peak region

In the build-up and Bragg peak region, positron-emitting fragments are produced via a combination of target fragmentation and projectile fragmentation. There is a rapid change in positron-emitting fragment yield with respect to depth, especially since different positron-emitting fragments stop at different distances from their point of production.

NMSE and Pearson CCC results between simulation and experimental total positron annihilation profiles in the build-up and Bragg peak region are summarised in Tables 4 and 5, respectively, with corresponding figures shown in Supplementary Material Section 2.

Table 4. Number of test cases for which each Geant4 version/model combination achieved the lowest or equal-lowest NMSE in the **build-up and Bragg peak region**. **Bold text** denotes the version/model achieving the highest (or equal-highest) number of best results for each combination of ion/energy/target.

Version	**Total**			**^{11}C**			**^{10}C**			**^{15}O**		
	BIC	**QMD**	**INCL**	**BIC**	**QMD**	**INCL**	**BIC**	**QMD**	**INCL**	**BIC**	**QMD**	**INCL**
10	4	0	0	11	6	3	0	0	2	3	2	0
10.1	5	0	0	11	6	3	0	0	2	3	2	0
10.2	**11**	1	0	**14**	6	3	0	0	1	**5**	1	0
10.3	1	0	0	5	0	0	4	1	2	3	0	0
10.4	1	0	0	1	0	0	3	2	2	4	0	0
10.5	0	0	0	0	0	0	3	4	7	1	2	0
10.6	1	0	0	0	0	0	3	0	**9**	2	1	2
10.7	1	0	0	0	0	0	3	2	7	0	0	1
11	2	0	0	0	0	0	3	2	**9**	2	0	2
11.1	1	0	0	0	0	0	3	2	**9**	2	0	2

Table 5. Number of test cases for which each Geant4 version/model combination achieved a CCC greater than 0.95 in the **build-up and Bragg peak region**. **Bold text** denotes the version/model achieving the highest number of best results for each combination of ion/energy/target.

Version	**Total**			**^{11}C**			**^{10}C**			**^{15}O**		
	BIC	**QMD**	**INCL**	**BIC**	**QMD**	**INCL**	**BIC**	**QMD**	**INCL**	**BIC**	**QMD**	**INCL**
10	9	4	3	3	3	2	4	3	**6**	3	2	3
10.1	9	4	4	3	3	3	4	3	**6**	3	2	3
10.2	**10**	8	4	6	6	3	5	4	5	3	3	3
10.3	9	6	6	6	5	3	2	2	2	**4**	3	3
10.4	6	8	7	6	5	3	1	1	3	**4**	3	3
10.5	8	8	6	6	5	4	2	2	3	3	3	**4**
10.6	9	9	6	6	**7**	6	2	1	4	**4**	3	**4**
10.7	6	6	4	3	5	2	2	1	4	3	2	3
11	8	8	6	4	5	4	3	2	4	3	3	3
11.1	6	5	6	2	3	4	3	2	4	3	3	3

In the build-up and Bragg peak region, according to the NMSE metric, total positron yield is most accurately predicted by the BIC model in Geant4 version 10.2, being (equal) best in 11/15 cases. This is much higher than the next-best combinations (10.1/BIC with 5/11 cases followed by 10/BIC with 4/11). Similar results are observed for ^{11}C yield, with 10.2/BIC achieving (equal) best performance in 14/15 cases, and 10/BIC and 10.1/BIC each achieving (equal) best results in 11/15 cases; QMD also performs reasonably well in this case with 10, 10.1 and 10.2 achieving wins in 6/15 cases. For ^{10}C, 10.6/INCL, 11/INCL and 11.1/INCL are the best performers (each winning in 9/15 cases). Finally, for ^{15}O, 10.2/BIC is the best-performing model with 5/12 wins, followed by 10.4/BIC with 4 wins.

Table 6. Differences between the depths of the maximum positron annihilation yield in experimental and simulation results. Each voxel has a width of 1.5 mm; the maximum error is always in multiples of 1.5 mm increments.

Version	**BIC**			**QMD**			**INCL**		
	μ **(mm)**	σ **(mm)**	**max (mm)**	μ **(mm)**	σ **(mm)**	**max (mm)**	μ **(mm)**	σ **(mm)**	**max (mm)**
10	1	1.85	6	-0.20	1.69	3	1.10	3.82	10.50
10.1	1	1.85	6	-0.20	1.69	3	1	3.91	10.50
10.2	0.60	1.37	3	-0.60	1.58	-3	0.60	3.39	9
10.3	1.30	1.78	6	-0.30	1.41	-3	2.30	4.12	9
10.4	1.60	1.55	6	-0.10	1.44	-3	4	4.45	10.50
10.5	0.69	0.99	3	0.60	2.03	6	2.20	4.24	9
10.6	0.60	1.37	3	-0.10	1.55	3	0.10	2.50	7.50
10.7	1.20	1.72	4.50	0.60	1.95	4.50	0.70	3.40	10.50
11	1.10	1.65	4.50	0.60	1.95	4.50	0.60	3.09	9
11.1	0.30	1.62	3	-0.30	1.62	3	0	2.78	7.50

Using the Pearson CCC metric, the best-performing version/model combinations for overall positron yield are 10.2/BIC (10/15 cases), followed by 10/BIC, 10.1/BIC, 10.3/BIC, 10.6/BIC and 10.6/QMD (9/15 cases). Generally, BIC performed very well, with all Geant4 versions achieving (equal) best performance in at least 6 cases. ^{11}C yield was best predicted by 10.6/QMD (7/15 cases) however many version/model combinations did well here also, with 10.2/BIC, 10.2/QMD, 10.3/BIC, 10.4/BIC, 10.5/BIC, 10.6/BIC and 10.6/INCL all achieving 6/15 wins. ^{10}C yield was best predicted by 10/INCL and 10.1 INCL (6/15 cases), closely followed by 10.2/BIC and 10.2/INCL which won in 5/15 cases. The best-performing version/model combinations for ^{15}O yield were 10.3/BIC, 10.4/BIC, 10.5/INCL, 10.6/BIC and 10.6/INCL with 4/15 wins each, and all other version/model combinations achieving 2 or 3 wins.

Table 6 lists difference between the experimental and simulation positron peak, while Table 7 lists the difference between the 50% fall off point for the experimental and simulated positron peak.

The smallest differences between experimental and simulation-based depth of maximum positron annihilation were obtained with Geant4 10.4/QMD ($\mu = -0.1$ mm; max = -3 mm) and 10.6/QMD ($\mu = -0.1$ mm; max = +3 mm). While a smaller mean value was obtained with 11.1/INCL, the maximum value and standard deviation were much larger (+7.5 mm and 2.78 mm) compared to 10.4/QMD and 10.6/QMD. Differences in the depth of the distal 50%-of-peak point were much smaller; the best estimates were obtained with 10.2/QMD ($\mu = 0$ mm; max = +1.5 mm), 11.1/BIC ($\mu = 0$ mm; max = -3 mm) and 11/INCL ($\mu = 0$ mm; max = +3 mm).

Table 7. Differences between the distal depths at which the positron annihilation yield has decreased to 50% of the peak value in experimental and simulation results. Each voxel has a width of 1.5 mm; the maximum error is always in multiples of 1.5 mm increments.

Version	**BIC**			**QMD**			**INCL**		
	μ **(mm)**	σ **(mm)**	**max (mm)**	μ **(mm)**	σ **(mm)**	**max (mm)**	μ **(mm)**	σ **(mm)**	**max (mm)**
10	0.70	1.49	3	0.30	1.41	3	-0.20	1.49	-3
10.1	0.70	1.49	3	0.30	1.41	3	-0.20	1.49	-3
10.2	0.30	1.01	1.50	0	1.13	1.50	-0.50	1.22	-3
10.3	0.50	1.09	3	0.20	1.11	1.50	-0.20	1.37	-3
10.4	0.60	1.11	3	0.30	1.01	1.50	0.10	1.20	-3
10.5	0.35	0.90	1.50	0.20	1.11	1.50	-0.50	1.22	-3
10.6	0.40	0.89	1.50	0.20	1.11	1.50	-0.40	1.33	-3
10.7	1	1.46	3	0.70	1.59	3	-0.10	1.65	3
11	1	1.46	3	0.70	1.59	3	0	1.60	3
11.1	0	1.60	-3	-0.10	1.44	-3	-0.60	1.68	-3

3.3. Tail region

In the tail region, positron-emitting radionuclides are primarily produced through fragmentation of the target material caused by light fragments created upstream from the primary beam. As such, the production of positron-emitting fragments in the tail region is highly dependent on fragmentation and scattering cross sections upstream. Therefore, the yield of positron annihilation is not expected to rapidly change across this region compared to the build-up and Bragg peak region.

NMSE and Pearson CCC results between simulation and experimental total positron annihilation profiles in the tail region are summarised in Tables 8 and 9, respectively, with corresponding figures shown in Supplementary Material Section 3.

Using the NMSE metric, 10.2/BIC was the best-performing version/model combination for overall positron yield (12/15 cases), with 10.2/QMD being the second-best (10/15). Results were similar for ^{11}C yield, with the best version/model combinations being 10.2/QMD (12/15 cases) and 10.2/BIC (11/15). For ^{10}C, the most wins were obtained by 10.6/INCL and 11/INCL (6/15 cases) followed by 10.6/BIC, 10.7/INCL and 11.1/INCL (5/15 cases). Finally, for ^{15}O, the best results were obtained with 10.6/QMD (10/12 cases) followed by 10.2/BIC and 10.2/QMD (7/12 cases).

The Pearson CCC results in the tail region were all very similar across Geant4 versions, with only a few wins separating the best and worst-performing version/model combinations in most instances. All version/models exceeded the threshold of 0.95 for a clear majority of cases for total positron yield as well as ^{11}C and ^{15}O production. For total positron annihilation yield, 10/BIC, 10.1/BIC, 10.2/BIC, 10.6/INCL, 10.7/BIC, 10.7/INCL, 11/BIC, 11/INCL, 11.1/BIC and 11.1/INCL all exceeded the target threshold for 12/15 cases. Even the worst-performing version/model combinations still

Table 8. Number of test cases for which each Geant4 version/model combination achieved the lowest or equal-lowest NMSE in the **tail region**. **Bold text** denotes the version/model achieving the highest (or equal-highest) number of best results for each combination of ion/energy/target.

Version	**Total**			**^{11}C**			**^{10}C**			**^{15}O**		
	BIC	**QMD**	**INCL**	**BIC**	**QMD**	**INCL**	**BIC**	**QMD**	**INCL**	**BIC**	**QMD**	**INCL**
10	3	2	2	5	9	4	2	2	1	4	4	4
10.1	4	2	2	6	9	4	1	1	2	4	4	4
10.2	**12**	10	2	11	**12**	4	1	1	3	7	7	4
10.3	1	1	1	1	1	1	2	4	0	3	4	4
10.4	1	1	1	1	1	1	2	3	0	3	4	3
10.5	1	1	1	1	1	1	3	4	2	2	2	2
10.6	1	2	1	1	1	1	5	3	**6**	4	**10**	4
10.7	1	1	1	1	1	1	3	3	5	3	4	3
11	1	1	1	1	1	1	4	3	**6**	4	4	3
11.1	1	1	1	1	1	1	4	3	5	4	4	3

Table 9. Number of test cases for which each Geant4 version/model combination achieved a CCC greater than 0.95 in the **tail region**. **Bold text** denotes the version/model achieving the highest number of best results for each combination of ion/energy/target.

Version	**Total**			**^{11}C**			**^{10}C**			**^{15}O**		
	BIC	**QMD**	**INCL**	**BIC**	**QMD**	**INCL**	**BIC**	**QMD**	**INCL**	**BIC**	**QMD**	**INCL**
10	**12**	11	11	10	11	11	**4**	3	**4**	8	7	**9**
10.1	**12**	11	11	11	11	11	**4**	**4**	**4**	8	7	**9**
10.2	**12**	11	11	10	11	11	**4**	**4**	**4**	8	7	7
10.3	11	11	10	11	11	11	2	2	1	8	7	8
10.4	10	10	10	9	11	10	2	2	0	7	7	8
10.5	11	11	11	11	11	11	2	2	3	7	7	8
10.6	11	11	**12**	11	11	11	2	3	3	7	7	**9**
10.7	**12**	11	**12**	11	11	12	3	3	3	7	7	**9**
11	**12**	11	**12**	11	11	**13**	3	3	3	8	8	**9**
11.1	**12**	11	**12**	11	11	**13**	3	3	3	8	8	**9**

exceeded the threshold in 10/15 cases. For ^{11}C yield, 11/INCL and 11.1/INCL reached the threshold in 13/15 cases (with the worst-performing combination scoring 9/15 wins). Fewer wins were seen with ^{10}C; the best results were obtained with 10/BIC, 10/INCL, 10.1/BIC, 10.1/QMD, 10.1/INCL, 10.2/BIC, 10.2/QMD and 10.2/INCL (4/15 cases). Finally, ^{15}O yield was best predicted by 10/INCL, 10.1/INCL, 10.6/INCL, 10.7/INCL, 11/INCL and 11.1/INCL (9/12 cases) - again, in this case, even the worst-performing version/model combinations exceeded the threshold in 7/12 cases.

3.4. Overall recommendation

The accuracy of Geant4's hadronic inelastic physics models (BIC, QMD and INCL) in predicting both total positron annihilation yield and individual positron-emitting fragment production is not consistent between different versions of Geant4; furthermore, later releases do not necessarily provide a more accurate prediction of experimental observations than preceding versions. In some cases, NMSE and Pearson CCC yielded conflicting results, due to the different features of the respective profiles which are emphasised by each metric (NMSE quantifying the overall average squared differences between the profiles while Pearson CCC quantifying the degree of linear dependence, independent of relative or absolute magnitude).

In the entrance region, BIC was clearly the best-performing model, with the best choice of Geant4 version depending on the particular metric and fragmentation product of interest. NMSE results generally favoured 10-10.4/BIC (especially 10.2/BIC), except for ^{10}C yield, which was better predicted by 10.3+/INCL. Pearson CCC performance did not strongly favour any particular version/model combination, with at most 1/3 of test cases achieving the target CCC threshold of 0.95 for any version/model.

In the build-up and Bragg peak region and tail region, the results are more conclusive. The NMSE metric conclusively shows that version 10.2/BIC is the best choice for total positron yield as well as ^{11}C and ^{15}O yield, while 10.5-11.1/INCL performed the best for ^{10}C. Pearson CCC results are more mixed, but again, 10.2/BIC gives the best results for total positron annihilation yield, with most versions of Geant4 with BIC performing well. 10.6/QMD performed the best for ^{11}C, 10/INCL and 10.1/INCL performed the best for ^{10}C, and there was no clear winner for ^{15}O.

Using the depth-of-maximum-yield metric, the smallest mean differences were obtained with 10.4/QMD and 10.6/QMD. These versions/models also achieved the equal-smallest maximum difference (-3 mm and +3 mm, respectively). Across all versions of Geant4, QMD demonstrated the best overall accuracy (lowest average mean difference in peak depth) and highest precision (lowest average standard deviation). INCL was the worst-performing model across all versions, with much larger maximum differences, and a consistent underestimation of depth of maximum yield across Geant4 versions, with the exception of version 11.1 (which, despite a mean difference of 0, exhibited a large standard deviation and maximum value). Standard deviations obtained using INCL were generally around double those of QMD and BIC. BIC also showed a consistent underestimation in depth of maximum yield, although the maximum differences were much smaller than for INCL. For context, the difference between the depth of the positron annihilation peak and the Bragg peak with monoenergetic ion beams is of the order of -5.6±0.8 mm for ^{12}C and -6.6±0.8 mm for ^{16}O [13, 14, 15].

Results were generally better for the the distal depth at 50% of peak metric. In this case, 10.2/QMD, 11.1/BIC and 11/INCL all achieved a mean of zero, with 10.2/QMD also having the equal-lowest maximum value of 1.5 mm (a depth difference of one voxel). QMD's maximal values were slightly smaller overall compared to BIC, and INCL's were the largest at ±3 mm for all versions. INCL tended to consistently overestimate the depth of this point, with both mean and maximum differences being negative in most cases. BIC and QMD both tended towards underestimating the 50%-of-peak depth, with the exception of version 11.1 (negative maxima for both, and means of 0 and -0.1 mm, respectively). Standard deviations were quite small for all versions and models (with the maximum standard deviation being 1.68 mm, for 11.1/INCL).

Finally, in the tail region, Geant4 10.2 with BIC and QMD again provided the best prediction of total positron and ^{11}C yield in terms of NMSE, while 10.6/INCL performed the best for ^{10}C and ^{15}O. All version/model combinations performed well for total positron annihilation, ^{11}C and ^{15}O yield according to the Pearson CCC metric, while no version/model performed especially well for ^{10}C.

Across all regions, ion species, beam energies, and target materials evaluated, the combination of Geant4 version 10.2 and BIC is best able to reproduce experimental results as evaluated using the NMSE and Pearson CCC metrics - especially in the build-up and Bragg peak region and tail region. Since the build-up and Bragg peak region is the location where (1) the majority of the dose resulting from carbon or oxygen ion beam irradiation in heavy ion therapy is deposited, and (2) where the strongest positron annihilation signal is observed, the results in this region are the most relevant to PET image-based QA simulation work. Version 10.2 also provided the best estimate of the depth of the distal depth at which positron yield decreased to 50% of peak, although this was obtained with QMD rather than BIC; the most accurate estimate of the depth of the peak itself was also achieved with QMD, but with Geant4 versions 10.4 and 10.6. As QMD exhibited the best accuracy and precision across Geant4 versions, it is the recommended model if the depth of the yield peak is critical.

The BIC model implemented in Geant4 version 10.5 suffered from a run-time stability error which resulted in it being unable to simulate all test scenarios; therefore, we recommend that this version/model combination should be avoided for future studies.

In the evaluation of individual positron-emitting fragment yield profiles, predictions of ^{10}C distribution were generally the least accurate in terms of both the NMSE and Pearson CCC. Interestingly, the INCL model often performed the best for prediction of ^{10}C fragment yield, although it rarely performed the best for total positron annihilation and ^{11}C or ^{15}O. Therefore, INCL should be considered for studies focusing on ^{10}C fragmentation, with the caveat that range estimation will be less accurate with this model.

Not all models met or exceeded the set threshold of 0.95 for the Pearson CCC metric. This means that in these cases, the shape of the predicted positron distribution differs significantly from the experimental measurements. This is of particular concern if these models are to be used for dose estimation using a deconvolution approach [2, 3] or for the training of machine learning models for feature extraction [16].

One may reasonably ask why the performance of the fragmentation models in Geant4 has not continued to steadily improve with each release, and in fact has regressed at times. Positron-emitting isotope production channels represent only a fraction of all possible reaction outcomes, so it may be the case that by improving results for one subset of reaction processes, the positron-emitting nuclide production cross sections became worse. Another possible reason is the implementation of different numbers of de-excitation channels in the Fermi Break-Up model in different versions of Geant4. Unfortunately, to date, no detailed investigation has been conducted into Geant4 to pin down the specific cause, and it is unknown at this stage if there are other contributing factors as well. In order to more strictly monitor the impact of the evolution of Geant4 in the results of a simulation application of interest, the Geant4 developers are developing an automated benchmarking system for medical applications in Geant4 (the G4-Med project) which should help to document the reasons behind different results when using different Geant4 releases with higher granularity [17].

In the next release of Geant4, 11.2, a new quantum molecular dynamics model, "Light Ion QMD", will be be introduced‖ with a specific focus on hadron therapy [18]. In future work, we will be collaborating with the developers of this model to compare its performance against the other models included in Geant4 11.2 with a focus on in vivo PET applications.

Finally, it is worth noting that current evaluations of fragmentation cross sections exhibit uncertainties exceeding 10%, which must be tightened in order to accurately model positron fragmentation, particularly in the case of complex fragmentation reactions such as the production of ^{10}C [19, 20]. These uncertainties are especially due to the effective cross-sections that are double-differential in angle and energy. Since these cross-sections provide a strong constraint on nucleus-nucleus reaction models, access to improved experimental measurements of these cross-sections is vital to constraining these models and improving their accuracy. This also impacts other Monte Carlo simulation platforms (such as FLUKA, MCNP and PHITS) which also rely on accurate cross section data (although notably PHITS uses a new version of this model, JQMD2, which tries to correct the main flaw of the QMD model, the drop in effective cross-sections at low angles [21]).

4. Conclusion

In this study, the accuracy with which Geant4 is able to predict the distribution of total positron annihilation yield and the distributions of individual positron-emitting fragmentation products (^{11}C, ^{10}C and ^{15}O) during carbon or oxygen ion therapy was compared to experimental data. Three different hadronic inelastic physics models - Binary Ion Cascade (BIC), Quantum Molecular Dynamics (QMD) and Liege Intranuclear Cascade model (INCL) were used with ten different versions of Geant4 - 10.0.p04, 10.1.p03, 10.2.p03, 10.3.p03, 10.4.p03, 10.5.p01, 10.6.p03, 10.7.p02, 11.0 and 11.1, in three different homogeneous phantoms. The simulated and experimental data were compared using two different metrics - normalised mean squared error and the Pearson cross-correlation coefficient. Additionally, the differences between the simulated

---

‖ Note: this model had not been included in Geant4 prior to the submission of this manuscript and experimental depth of maximum positron annihilation yield, as well as the distal point at which positron yield declines to 50% of the peak were evaluated. It was found that the accuracy of the hadronic inelastic physics models strongly depends on the version of Geant4 in which it was implemented, and newer versions of Geant4 were not always more accurate at predicting positron-emitting fragmentation compared to older versions. Furthermore, it was found that not all version/model combinations were able to satisfactorily predict the shape of positron annihilation or positron-emitting fragment distributions, even though they could provide a good estimation of the total positron annihilation yield and range. For future simulation studies of therapeutic irradiation using carbon or oxygen ion beams, it is recommended that Geant4 version 10.2 with the BIC model be used as it is currently the version/model combination best able to replicate the experimentally-observed total positron yield and the fragmentation product distributions, while the depth of the maximum positron yield and distal 50%-of-peak point were best predicted using the QMD model from Geant4 10.4, 10.6 (peak) and 10.2 (distal 50%-of-peak).

5. Acknowledgements

The authors would like to acknowledge the following organisations for providing access to their high-performance computing resources: the Multi-modal Australian Sciences Imaging and Visualisation Environment (MASSIVE) "M3" cluster and Australia's Nuclear Science and Technology Organisation (ANSTO) "Tesla" cluster. This research has been conducted with the support of the Australian government research training program scholarship. The authors acknowledge the scientific and technical assistance of the National Imaging Facility, a National Collaborative Research Infrastructure Strategy (NCRIS) capability at the Australian Nuclear Science and Technology Organisation, ANSTO.

References

[1] M. Durante, R. Orecchia, J. S. Loeffler, Charged-particle therapy in cancer: clinical uses and future perspectives, Nature Reviews Clinical Oncology 14 (8) (2017) 483–495. `doi:10.1038/nrclinonc.2017.30`.
URL `https://doi.org/10.1038/nrclinonc.2017.30`

[2] T. Hofmann, M. Pinto, A. Mohammadi, M. Nitta, F. Nishikido, Y. Iwao, H. Tashima, E. Yoshida, A. Chacon, M. Safavi-Naeini, A. Rosenfeld, T. Yamaya, K. Parodi, Dose reconstruction from PET images in carbon ion therapy: a deconvolution approach, Physics in Medicine & Biology 64 (2) (2019) 025011. `doi:10.1088/1361-6560/aaf676`.
URL `https://doi.org/10.1088/1361-6560/aaf676`

[3] T. Hofmann, A. Fochi, K. Parodi, M. Pinto, Prediction of positron emitter distributions for range monitoring in carbon ion therapy: an analytical approach, Physics in Medicine & Biology 64 (10) (2019) 105022. `doi:10.1088/1361-6560/ab17f9`.
URL `https://doi.org/10.1088/1361-6560/ab17f9`

[4] H. Rutherford, A. Chacon, A. Mohammadi, S. Takyu, H. Tashima, E. Yoshida, F. Nishikido, T. Hofmann, M. Pinto, D. R. Franklin, T. Yamaya, K. Parodi, A. B. Rosenfeld, S. Guatelli,

M. Safavi-Naeini, Dose quantification in carbon ion therapy using in-beam positron emission tomography, Physics in Medicine & Biology 65 (23) (2020) 235052. `doi:10.1088/1361-6560/abaa23`.
URL `https://doi.org/10.1088/1361-6560/abaa23`

[5] K. Parodi, J. C. Polf, In vivo range verification in particle therapy, Medical Physics 45 (11) (Nov. 2018). `doi:10.1002/mp.12960`.
URL `https://doi.org/10.1002/mp.12960`

[6] S. Agostinelli, J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. Barrand, F. Behner, L. Bellagamba, J. Boudreau, L. Broglia, A. Brunengo, H. Burkhardt, S. Chauvie, J. Chuma, R. Chytracek, G. Cooperman, G. Cosmo, P. Degtyarenko, A. Dell'Acqua, G. Depaola, D. Dietrich, R. Enami, A. Feliciello, C. Ferguson, H. Fesefeldt, G. Folger, F. Foppiano, A. Forti, S. Garelli, S. Giani, R. Giannitrapani, D. Gibin, J. G. Cadenas, I. González, G. G. Abril, G. Greeniaus, W. Greiner, V. Grichine, A. Grossheim, S. Guatelli, P. Gumplinger, R. Hamatsu, K. Hashimoto, H. Hasui, A. Heikkinen, A. Howard, V. Ivanchenko, A. Johnson, F. Jones, J. Kallenbach, N. Kanaya, M. Kawabata, Y. Kawabata, M. Kawaguti, S. Kelner, P. Kent, A. Kimura, T. Kodama, R. Kokoulin, M. Kossov, H. Kurashige, E. Lamanna, T. Lampén, V. Lara, V. Lefebure, F. Lei, M. Liendl, W. Lockman, F. Longo, S. Magni, M. Maire, E. Medernach, K. Minamimoto, P. M. de Freitas, Y. Morita, K. Murakami, M. Nagamatu, R. Nartallo, P. Nieminen, T. Nishimura, K. Ohtsubo, M. Okamura, S. O'Neale, Y. Oohata, K. Paech, J. Perl, A. Pfeiffer, M. Pia, F. Ranjard, A. Rybin, S. Sadilov, E. D. Salvo, G. Santin, T. Sasaki, N. Savvas, Y. Sawada, S. Scherer, S. Sei, V. Sirotenko, D. Smith, N. Starkov, H. Stoecker, J. Sulkimo, M. Takahata, S. Tanaka, E. Tcherniaev, E. S. Tehrani, M. Tropeano, P. Truscott, H. Uno, L. Urban, P. Urban, M. Verderi, A. Walkden, W. Wander, H. Weber, J. Wellisch, T. Wenaus, D. Williams, D. Wright, T. Yamada, H. Yoshida, D. Zschiesche, Geant4—a simulation toolkit, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 506 (3) (2003) 250–303. `doi:10.1016/s0168-9002(03)01368-8`.
URL `https://doi.org/10.1016/s0168-9002(03)01368-8`

[7] G. Collaboration, Physics reference manual for geant4, Tech. rep., CERN (2018).

[8] D. Mancusi, A. Boudard, J. Cugnon, J.-C. David, P. Kaitaniemi, S. Leray, Extension of the Liège intranuclear-cascade model to reactions induced by light nuclei, Phys. Rev. C 90 (2014) 054602. `doi:10.1103/PhysRevC.90.054602`.
URL `https://link.aps.org/doi/10.1103/PhysRevC.90.054602`

[9] A. Boudard, J. Cugnon, S. Leray, C. Volant, Intranuclear cascade model for a comprehensive description of spallation reaction data, Physical Review C 66 (4) (Oct. 2002). `doi:10.1103/physrevc.66.044615`.
URL `http://dx.doi.org/10.1103/PhysRevC.66.044615`

[10] A. Chacon, S. Guatelli, H. Rutherford, D. Bolst, A. Mohammadi, A. Ahmed, M. Nitta, F. Nishikido, Y. Iwao, H. Tashima, E. Yoshida, G. Akamatsu, S. Takyu, A. Kitagawa, T. Hofmann, M. Pinto, D. R. Franklin, K. Parodi, T. Yamaya, A. Rosenfeld, M. Safavi-Naeini, Comparative study of alternative geant4 hadronic ion inelastic physics models for prediction of positron-emitting radionuclide production in carbon and oxygen ion therapy, Physics in Medicine & Biology 64 (15) (2019) 155014. `doi:10.1088/1361-6560/ab2752`.
URL `https://doi.org/10.1088/1361-6560/ab2752`

[11] G. Akamatsu, H. Tashima, Y. Iwao, H. Wakizaka, T. Maeda, A. Mohammadi, S. Takyu, M. Nitta, F. Nishikido, H. Rutherford, A. Chacon, M. Safavi-Naeini, E. Yoshida, T. Yamaya, Performance evaluation of a whole-body prototype PET scanner with four-layer DOI detectors, Physics in Medicine and Biology (Apr. 2019). `doi:10.1088/1361-6560/ab18b2`.
URL `https://doi.org/10.1088/1361-6560/ab18b2`

[12] T. Swinscow, Statistics at Square One, Wiley, 2021. `doi:10.1002/9781119402350`.
URL `http://dx.doi.org/10.1002/9781119402350`

[13] A. Chacon, B. James, L. Tran, S. Guatelli, L. Chartier, D. Prokopvich, D. R. Franklin, A. Mohammadi, F. Nishikido, Y. Iwao, G. Akamatsu, S. Takyu, H. Tashima, T. Yamaya, K. Parodi, A. Rosenfeld, M. Safavi-Naeini, Experimental investigation of the characteristics of radioactive beams for heavy ion therapy, Medical Physics 47 (7) (2020) 3123–3132. `doi:10.1002/mp.14177`.
URL `http://dx.doi.org/10.1002/mp.14177`

[14] R. S. Augusto, A. Mohammadi, H. Tashima, E. Yoshida, T. Yamaya, A. Ferrari, K. Parodi, Experimental validation of the fluka monte carlo code for dose and β^+ -emitter predictions of radioactive ion beams, Physics in Medicine & Biology 63 (21) (2018) 215014. `doi:10.1088/1361-6560/aae431`.
URL `http://dx.doi.org/10.1088/1361-6560/aae431`

[15] A. Mohammadi, H. Tashima, Y. Iwao, S. Takyu, G. Akamatsu, F. Nishikido, E. Yoshida, A. Kitagawa, K. Parodi, T. Yamaya, Range verification of radioactive ion beams of ^{11}C and ^{15}O using in-beam PET imaging, Physics in Medicine & Biology 64 (14) (2019) 145014. `doi:10.1088/1361-6560/ab25ce`.
URL `https://doi.org/10.1088/1361-6560/ab25ce`

[16] H. Rutherford, R. S. Turai, A. Chacon, D. R. Franklin, A. Mohammadi, H. Tashima, T. Yamaya, K. Parodi, A. B. Rosenfeld, S. Guatelli, M. Safavi-Naeini, An inception network for positron emission tomography based dose estimation in carbon ion therapy, Physics in Medicine and Biology 67 (19) (2022) 194001. `doi:10.1088/1361-6560/ac88b2`.
URL `https://doi.org/10.1088/1361-6560/ac88b2`

[17] P. Arce, D. Bolst, M. Bordage, J. M. C. Brown, P. Cirrone, M. A. Cortés-Giraldo, D. Cutajar, G. Cuttone, L. Desorgher, P. Dondero, A. Dotti, B. Faddegon, C. Fedon, S. Guatelli, S. Incerti, V. Ivanchenko, D. Konstantinov, I. Kyriakou, G. Latyshev, A. Le, C. Mancini-Terracciano, M. Maire, A. Mantero, M. Novak, C. Omachi, L. Pandola, A. Perales, Y. Perrot, G. Petringa, J. M. Quesada, J. Ramos-Méndez, F. Romano, A. B. Rosenfeld, L. G. Sarmiento, D. Sakata, T. Sasaki, I. Sechopoulos, E. C. Simpson, T. Toshito, D. H. Wright, Report on G4-Med, a Geant4 benchmarking system for medical physics applications developed by the Geant4 Medical Simulation Benchmarking Group, Medical Physics 48 (1) (2020) 19–56. `doi:10.1002/mp.14226`.
URL `http://dx.doi.org/10.1002/mp.14226`

[18] Y.-h. Sato, D. Sakata, D. Bolst, E. C. Simpson, S. Guatelli, A. Haga, Development of a more accurate geant4 quantum molecular dynamics model for hadron therapy, Physics in Medicine and Biology 67 (22) (2022) 225001. `doi:10.1088/1361-6560/ac9a9a`.
URL `http://dx.doi.org/10.1088/1361-6560/ac9a9a`

[19] D. Bolst, G. A. Cirrone, G. Cuttone, G. Folger, S. Incerti, V. Ivanchenko, T. Koi, D. Mancusi, L. Pandola, F. Romano, A. B. Rosenfeld, S. Guatelli, Validation of geant4 fragmentation for heavy ion therapy, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 869 (2017) 68–75. `doi:10.1016/j.nima.2017.06.046`.
URL `https://doi.org/10.1016/j.nima.2017.06.046`

[20] M. Toppi, A. Sarti, A. Alexandrov, B. Alpat, G. Ambrosi, S. Argirò, R. A. Diaz, M. Barbanera, N. Bartosik, G. Battistoni, N. Belcari, S. Biondi, M. G. Bisogni, M. Bon, G. Bruni, P. Carra, F. Cavanna, P. Cerello, E. Ciarrocchi, A. Clozza, S. Colombi, G. D. Lellis, A. D. Gregorio, A. D. Guerra, M. D. Simoni, A. D. Crescenzo, B. D. Ruzza, M. Donetti, Y. Dong, M. Durante, V. Ferrero, E. Fiandrini, C. Finck, E. Fiorina, M. Fischetti, M. Francesconi, M. Franchini, G. Franciosini, G. Galati, L. Galli, G. Giraudo, R. Hetzel, E. Iarocci, M. Ionica, A. Iuliano, K. Kanxheri, A. Kraan, C. L. Tessa, M. Laurenza, A. Lauria, E. L. Torres, M. Marafini, M. Massa, C. Massimi, I. Mattei, A. Meneghetti, A. Mengarelli, R. Mirabelli, A. Moggi, M. Montesi, M. Morone, M. Morrocchi, S. Muraro, F. Murtas, A. Muscato, A. Pastore, N. Pastrone, V. Patera, F. Pennazio, F. Peverini, P. Placidi, M. Pullia, L. Ramello, C. Reidel, R. Ridolfi, V. Rosso, C. Sanelli, G. Sartorelli, O. Sato, S. Savazzi, L. Scavarda, A. Schiavi,

C. Schuy, E. Scifoni, A. Sciubba, A. Sécher, M. Selvi, L. Servoli, G. Silvestre, M. Sitta, R. Spighi, E. Spiriti, G. Sportelli, A. Stahl, S. Tomassini, F. Tommasino, V. Tioukov, G. Traini, A. Trigilio, S. Valle, M. Vanstalle, U. Weber, R. Zarrella, A. Zoccoli, M. Villa, Elemental fragmentation cross sections for a ^{16}O beam of 400 MeV/u kinetic energy interacting with a graphite target using the FOOT ΔE-TOF detectors, Frontiers in Physics 10 (Nov. 2022). `doi:10.3389/fphy.2022.979229`.
URL `https://doi.org/10.3389/fphy.2022.979229`

[21] T. Ogawa, T. Sato, S. Hashimoto, D. Satoh, S. Tsuda, K. Niita, Energy-dependent fragmentation cross sections of relativistic ^{12}, Physical Review C 92 (2) (Aug. 2015). `doi:10.1103/physrevc.92.024614`.
URL `http://dx.doi.org/10.1103/PhysRevC.92.024614`

6. Appendix

Table 10 lists the physics models which were used in the simulations.

Table 10. Hadronic physics processes and models used in all simulations.

Interaction	**Energy Range**	**Geant4 Model**
Radioactive Decay	All energies	G4RadioactiveDecayPhysics
Particle Decay	All energies	G4Decay
Hadron Elastic	0–100 TeV	G4HadronElasticPhysicsHP
Ion Inelastic	<100 MeV	Binary Light Ion Cascade
	100 MeV–10 GeV	BIC or QMD or INCL++
Neutron Capture	0–20 MeV	NeutronHPCapture
	>19.9 MeV	nRadCapture
Neutron Inelastic	0–20 MeV	NeutronHPInelastic
	>19.9 MeV	Binary Cascade
Proton Inelastic	990 eV–10 TeV	Binary Cascade