

EURADOS Report 2025-01

Oberschleißheim, July 2025



Participation of EURADOS WG13 in the AAPM TACTIC Grand Challenge: Contribution to TAC Curve Fitting in Nuclear Medicine Dosimetry

Vladimir Spielmann, Ziyi Hu, László Mázik,
Balázs Madas, Wei Bo Li

ISSN 2226-8057

ISBN 978-3-943701-39-5

DOI 10.12768/3gp7-6x92

Participation of EURADOS WG13 in the AAPM TACTIC Grand Challenge: Contribution to TAC Curve Fitting in Nuclear Medicine Dosimetry

Vladimir Spielmann¹, Ziyi Hu², László Mázik³,
Balázs Madas³, Wei Bo Li¹

¹ Federal Office for Radiation Protection BfS, Oberschleißheim, Germany

² Tsinghua University, Beijing, China

³ Institute for Energy Security and Environmental Safety HUN-REN Centre for Energy Research, Budapest, Hungary

ISSN 2226-8057

ISBN 978-3-943701-39-5

DOI 10.12768/3gp7-6x92

List of participants in the AAMP Grand Challenge (affiliations at the time of the exercise)

Vladimir Spielmann Lena Katzdobler Shamil Galvez Febles Wei Bo Li	Federal Office for Radiation Protection BfS, Oberschleißheim, Germany
Ziyi Hu Rui Qiu	Tsinghua University, Beijing, China
Carlos Gomes Ferreira Kuangyu Shi	Bern University, Bern, Switzerland
László Mázik Balázs Madas	Institute for Energy Security and Environmental Safety, HUN-REN Centre for Energy Research, Budapest, Hungary
Tomáš Vrba	Czech Technical University in Prague, Prag, Czech
David Broggio	RSN, Fontenay-aux-Roses, France
Dragoslav Nikezic	University of Kragujevac, Kragujevac, Serbia
Pedro Teles	Research Centre of Portuguese Cancer Institute (IPO-PORTO) and Faculty of Sciences of University of Porto, Porto, Portugal

Imprint

© EURADOS 2025

Issued by:

European Radiation Dosimetry Group e. V.
Ingolstädter Landstraße 1
85764 Oberschleißheim
Germany
office@eurados.org
www.eurados.org

The European Radiation Dosimetry Group e.V. is a non-profit organization promoting research and development and European cooperation in the field of the dosimetry of ionizing radiation. It is registered in the Register of Associations (Amtsgericht München, registry number VR 207982) and certified to be of non-profit character (Finanzamt München, notification from 2025-01-31).

Liability Disclaimer

No liability will be undertaken for completeness, editorial or technical mistakes, omissions as well as for correctness of the contents.

Table of contents

Table of contents.....	i
Abstract.....	iii
1. The AAPM Grand Challenge TACTIC.....	1
1.1. Background.....	1
1.2. Rules of competition.....	2
1.3. EURADOS participation in the grand challenge.....	2
2. Grand Challenge Phase 0.....	3
2.1. Introduction and motivation.....	3
2.2. Data.....	3
2.3. Curve fitting techniques/methods.....	4
2.3.1. SigmaPlot / OriginPro.....	9
2.3.2. MATLAB.....	9
2.3.3. SAAM II.....	9
2.3.3.1. Curve-fitting with SAAM II numerical module:.....	10
2.3.3.2. Compartmental model fitting with SAAM II compartmental module:.....	10
2.4. Goodness of fit.....	13
2.4.1. Coefficient of determination.....	13
2.4.2. Correlation matrix.....	13
2.4.3. Coefficients of variation (CV) of the fitted parameters.....	14
2.4.4. Statistical methods for model selection.....	14
2.5. Error models in the fitting.....	16
2.6. Results.....	16
2.6.1. Blood.....	16
2.6.2. Kidneys.....	20
2.6.3. Tumor.....	23
2.7. Discussions.....	24
2.8. Conclusions.....	26
3. Grand Challenge Phase 1.....	27
3.1. Introduction and motivation.....	27
3.2. Data.....	27
3.3. Consider value at time point 0 for blood data.....	28
3.4. Results.....	28
3.4.1. Blood.....	28

3.4.2.	Kidneys.....	30
3.4.3.	Tumor.....	32
3.5.	Discussion.....	34
3.6.	Conclusion	35
4.	Grand Challenge Phase 2	36
4.1.	Introduction and motivation	36
4.2.	Data	36
4.3.	Discussion on individual and population model	37
4.3.1.	PBMS (Population-based model selection).....	37
4.3.2.	SP-PBMS (Shared-parameter population-based model selection).....	38
4.3.3.	NLMEMS (Non-Linear Mixed-Effects model selection)	38
4.3.4.	NLME-PBMS (Population-Based Model Selection within the framework of the Non-Linear Mixed-Effects).....	39
4.4.	Results.....	40
4.4.1.	Blood	40
4.4.2.	Kidneys.....	42
4.4.3.	Tumor.....	45
4.5.	Discussion.....	48
4.6.	Conclusion	51
5.	Lessons learned from the TACTIC Grand Challenge.....	52
5.1.	Introduction	52
5.2.	Lessons we learned and recommendations.....	52
5.2.1.	Choose the right temporal frequency.....	52
5.2.2.	Consider a priori knowledge.....	53
5.2.2.1.	Consider values at time point zero.....	53
5.2.2.2.	Use an appropriate set of fit functions. Include known function parameters	53
5.2.2.3.	Implement error models.....	54
5.2.3.	Use appropriate integration methods	54
5.2.4.	Use appropriate model selection approach	55
5.2.5.	Consider goodness of fit	55
5.3.	Conclusion	56
6.	References.....	58

Abstract

On January 19, 2023, AAPM opened the TACTIC Grand Challenge for pre-registration. The TACTIC Grand Challenge organized the competition into three distinct phases, Phase 0, Phase 1 and Phase 2 with defined time frames, each designed to address key challenges in the field of nuclear medicine and radiopharmaceutical therapy. The three phases were structured as follows. Phase 0 was an optional warm-up exercise. Participants as team might select between 1 and 5 time points from a synthetic predefined dataset of 140 time points < 140 hours post injection of [^{177}Lu]Lu-PSMA. Participants should fit the time-activity curve (TAC) function that best represents the given data points, calculate the time-integrated activity (TIA) values based on each fitted TAC function. In Phase 1, participants might download the same predefined set of time-activity data corresponding to a clinical data acquisition scheme, consisting of 5 time points for 3 organs (blood, tumor and kidneys) for 5 patients (other than those used in Phase 0). Participants should perform fitting the TAC function that best represents the given data points, calculate the TIAC values based on each fitted TAC function. In Phase 1, individual curve-fitting techniques were mainly applied. In Phase 2, participants as teams might download the same predefined set of time-activity data corresponding to a clinical data acquisition scheme, consisting of 5 time points for 3 organs (blood, tumor and kidneys) for 15 additional patients (other than those used in Phase 1). Participants should fit the TAC functions that best represent the given data points, however applying the state-of-the-art population-based curve-fitting techniques. EURADOS Working Group (WG) 13 registered as one of the teams with the name EURADOS and participated in all three phases of the challenges. In Phase 0, EURADOS team achieved the best performer position. In Phase 1, EURADOS has mainly learned the techniques, how to treat the data at the time point zero. In Phase 2, EURADOS reached the third position with the best place in the competition in curve fitting for kidneys. Throughout the participation and performance in the challenges, the team EURADOS gained experience in many theoretical and technical methods for curve fitting in nuclear medicine, especially in the radiopharmaceutical therapy. These methods can be combined with individual and population compartmental modelling in the WG13 project and tasks.

1. The AAPM Grand Challenge TACTIC

1.1. Background

Calculation of internal dose are generally performed in 2 steps: description of the distribution in the body (biokinetic models) and description of the absorption of emitted radiation energy in the body's organs, i.e. phantoms and radiation transport (dosimetric models). Dosimetry in nuclear medicine has been well established by the Medical Internal Radiation Dose (MIRD) Committee of the Society of Nuclear Medicine and Molecular Imaging (SNMMI) and the International Commission on Radiological Protection (ICRP) (Pamphlet No. 1; MIRD Primer; ICRP 128, Bolch et al. 2009). Since then, more emerging radiopharmaceuticals have been developed for therapy, for example ^{223}Ra -dichloride, which was approved by Food and Drug Administration (FDA) on May 15, 2013 (FDA announcement). Later, additional radiopharmaceuticals, such as ^{177}Lu -DOTATATE (Abuqbeitah et al. 2018) and [^{177}Lu]Lu-PSMA-617 (Kabasakal et al. 2017) were approved by FDA and European Medical Agency (EMA). In 2013, the European Commission (EC) had already formalized its Directive 2013/59/Euratom, laying down basic safety standards (BSS) for radiation protection (EC directive 2013/59/Euratom) and this Directive is valid for all forms of radiotherapy including nuclear medicine therapy. Especially, EANM provided a guidance on how to interpret the Directive statements on the article 56 for nuclear medicine treatments (Konijnenberg 2021). Therefore, patient dosimetry for radiopharmaceutical therapy is, in contrast to radiopharmaceutical diagnostics, a more complex task and should be done in patient-specific manner. In addition to the EC Directive, the International Commission on Radiation Measurement and Unit (ICRU) published its Report 96 for dosimetry-guided radiopharmaceutical therapy (RPT) (ICRU 96), and European Association on Nuclear Medicine (EANM) published guidance document for dosimetry for first-in-human studies and early phase clinical trials (Stokke et al 2024).

Two patient-specific factors are employed in the dose assessment in nuclear medicine therapy. The first one, so called S-factor, is connected to the volumes of pharmaceutical accumulation tissues (source regions) and irradiation of the tissues or organs (target regions). Target regions cover real therapy targets, likewise healthy tissues, so-called organs at risk (OAR), which may be affected collaterally due to therapy. The regions had to have patient-specific topology, thus individual S-factors need to be calculated by means of Monte Carlo transport method or other techniques.

The second factor is related to the patient-specific individual biokinetics.

Generally, the time course of the radiopharmaceutical distribution, namely the time-activity curve (TAC), in each relevant region should be evaluated individually, so that the time-integrated activity coefficients (TIACs) can be calculated for each patient. Results from the SNMMI ^{177}Lu dosimetry challenge 2021 (Uribe et al. 2021) have demonstrated that RPT dosimetry is extremely user- and software-dependent. Although the TIAC values are crucial in clinical practice, many factors in the RPT dosimetry methodology of individual steps within the workflow, are not in consensus and influence the estimated absorbed dose to the patients, and therefore need to be standardized. In the beginning of 2023, AAPM was sponsoring a challenge on the calculation of TIACs in radiopharmaceutical therapy dosimetry. This AAPM Grand Challenge was named TACTIC challenge and focuses on the important step for RPT dosimetry, i.e. TAC and time-integrated activity (TIA) determination. The EURADOS WG13 "Dosimetry in Nuclear Medicine" organized an EURADOS team and participated in this challenge. This report summaries the participation activity and presents the

results. Furthermore, the experiences gathered during the activities are recommended for implementation in the biokinetic modelling and patient dosimetry in WG13.

1.2. Rules of competition

- The TACTIC Challenge consists of 3 phases: Phase 0 (15th February – 15th March), Phase 1 (16th March – 13th April) and Phase 2 (14th April – 15th May).
- Each participant must register an account through the challenge website for participation.
- Each participant team can compose one person or several persons.
- One individual person, who participated in one team, cannot participate in the competition individually. In Phase 0, each participant was asked to select five patient data and perform the curve fitting and submit the results for these five patients. However, the results of Phase 0 will not be considered as ranking.
- In Phase 1, participants will download the predefined dataset of five new patients and perform the curve fitting and submit the results for these five patients.
- In Phase 2, participants will download the predefined additional dataset from 15 patients, however, this dataset of 15 patients should be used with the dataset of the five patients in Phase 1 applying population fitting models to improve curve fitting for the five patients in Phase 1 and submit the improved fitting data for these 5 patients of Phase 1.
- All valid submissions from Phases 1 and 2 will be scored by the Challenge Team, who will compare results to the ground truth. The participants with the best score in Phase 1 and Phase 2 will be announced as the two winners.

1.3. EURADOS participation in the grand challenge

WG13 “Dosimetry in Nuclear Medicine” established EURADOS team and participated in all the three phases of the challenges. The participants of the EURADOS team are listed in the following with affiliations.

Name	Affiliation
David Broggio	IRSN, Fontenay-aux-Roses, France
Carlos Gomes Ferreira	Bern University, Bern, Switzerland
Shamil Galvez Febles	BfS-Federal Office for Radiation Protection, Oberschleißheim, Germany
Ziyi Hu	Tsinghua University, Beijing, China
Lena Katzdobler	BfS-Federal Office for Radiation Protection, Oberschleißheim, Germany
Wei Bo Li	BfS-Federal Office for Radiation Protection, Oberschleißheim, Germany
Balázs Madas	EK, Budapest, Hungary
László Mázik	EK, Budapest, Hungary
Dragoslav Nikezic	University of Kragujevac, Kragujevac, Serbia
Rui Qiu	Tsinghua University, Beijing, China
Kuangyu Shi	Bern University, Bern, Switzerland
Vladimir Spielmann	BfS-Federal Office for Radiation Protection, Oberschleißheim, Germany
Pedro Teles	University of Porto, Porto, Portugal
Tomáš Vrba	Czech Technical University in Prague, Prague, Czech

2. Grand Challenge Phase 0

2.1. Introduction and motivation

Phase 0 was an optional warm-up exercise phase.

The participation teams have to select between 1 and 5 time points from a synthetic predefined dataset from 1 to 140 hours post injection of [^{177}Lu]Lu-PSMA.

For chosen time points the organizers provided the participants with the following data for five patients:

- blood activity normalized to administered activity, in percent (%)
- kidneys activity normalized to administered activity, in percent (%)
- tumor activity normalized to administered activity and its volume, in percent (%)

The participants had to perform the following tasks:

- Fit the time-activity curve function that best represents the given data points
- Calculate the TIAs based on fitted TAC function (since the data provided by the organizers were normalized to administered activity, the final calculated values were TIACs)
- Complete the submission template file, to answer the following questions:
 - which formula of the fit functions were used
 - which parameters were used for all patients within the organ
 - whether a data uncertainty model was incorporated into the fit
 - how the goodness of fits was checked
 - whether the population information was considered in the used algorithm
 - if any additional information is needed to reproduce the results
- Submit the results back to the Challenge webpage

All valid submissions from Phase 0 were scored and ranked by the Challenge Team according to minimum root-mean-square error (RMSE) and maximum deviation of the results from the ground truth TIAC values of the synthetic data.

2.2. Data

As a general rule of thumb, at least as many data points should be obtained as the number of initially unknown variables in the mathematical curve-fitting function. Each exponential term in a multiexponential curve-fitting function requires three data points in order to be adequately characterized (Siegel et al. 1999). At least for kidneys and tumor a bi-exponential function might be expected. For this reason, and as the organizers planned to provide participants with five data points in Phases 1 and 2, it was decided to request five data points in Phase 0 as well.

The time points requested had to be the same for all organs and patients.

Selection of optimal time points for curve fitting depends on uptake and retention pattern of the radiopharmaceutical in each organ. The physical half-life or rather the effective half-time of the radiopharmaceutical compounds plays a decisive role in selecting an appropriate temporal frequency. When the distribution of selected time points is appropriately spread over the time interval, in our case between 0 and 140 h, then the mathematically simulated curve would be an accurate representation of the TAC.

If the radiopharmaceutical shows rapid uptake or washout, two or three time points at early times p.a. may need. Usually, the washout can be divided in two stages, early rapid clearance and long-term retention. Similarly, two or three data points may be required during the long-term retention phase to adequately characterize the late time course of the radiopharmaceutical.

[¹⁷⁷Lu]Lu-PSMA is a widely used radiopharmaceutical compound for targeted therapy in prostate cancer. Many publications regarding ¹⁷⁷Lu peptide receptor radionuclide therapy (PRRT) illustrated the importance of an appropriate sampling schedule for TAC estimation in personalized dosimetry (Ivashchenko et al. 2024). The uptake and retention pattern of [¹⁷⁷Lu]Lu-PSMA in different organs is described in numerous studies. In the pharmacokinetic analyses, it was shown that after injection, [¹⁷⁷Lu]Lu-PSMA-617 was distributed to the different organs within 2.5 h (Novartis Pharmaceuticals Corporation 2025). Clearance of the radiopharmaceutical from the body in the first few hours after administration is fast and then slowed down, resulting in the calculated effective half-lives of 1.7 ± 0.8 h and 41.1 ± 9.3 h, respectively (Hennrich and Eder 2022). There is substantial work within the literature exploring optimum timing regimens for therapies with [¹⁷⁷Lu]Lu-PSMA, particularly focusing on the pharmacokinetics of the kidneys. It is known that normal-tissue clearance is largely completed within the first 48 h after administration, whereas tumor retention is more closely reflected by the physical half-life of ¹⁷⁷Lu ($T_{1/2} = 6.6$ d) (Kratochwil et al. 2016; Jackson et al. 2020).

Usually, to characterize the long-term retention, time points equal to multiples of the effective half-time are needed. For bi-exponential functions, 0.1, 0.3, 1, 3 and 5 times of the effective half-life could be considered (Ivashchenko et al. 2024). The mean value determined for the effective half-life is approximately 41 h in [¹⁷⁷Lu]Lu-PSMA-617-targeted therapies. This fact would result in the following time points: 4, 12, 40, 120 and 200 h. In a study about influence of sampling schedules on [¹⁷⁷Lu]Lu-PSMA dosimetry (Rinscheid et al. 2020) it was shown that a late time point at 192 h p.i., is necessary for accurate and precise determination of tumor TIAC. As the latest time point, we could request, was 140 h, we selected this time point. Also, the optimal sampling schedules with two additional time points were identified to be 3-4 and 96-100 h in this study. So, we selected two time points in these intervals and additionally two other time points to cover the uptake stage and the interval between 4 h and 96 h.

According to this information about uptake, clearance and retention of [¹⁷⁷Lu]Lu-PSMA in blood, kidneys and tumor, it was finally considered useful to request the time points at 1, 4, 20, 96, and 140 h.

2.3. Curve fitting techniques/methods

Figure 1 shows the percent (%) activity data for kidneys, tumor and blood from five separate patients at requested time points provided by the organizers. The data points are connected by a dashed line to better distinguish individual patient data from each other. No information about data uncertainty was provided.

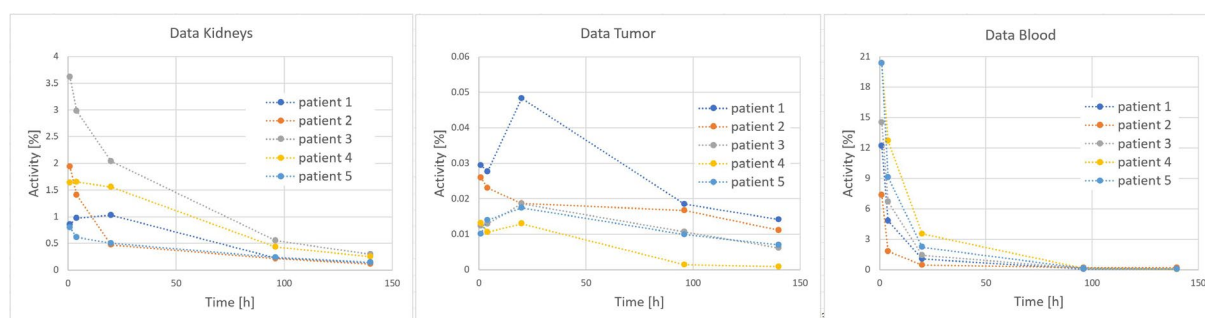


Figure 1. Percent (%) activity data for kidneys, tumor, and blood from five separate patients.

Since only a few measured time points are available, the complete pharmacokinetics for all points in time must be determined by interpolation between the measuring points and also extrapolation for times outside the measuring points. This interpolation can be performed using empirical methods, analytical methods or compartmental models.

An example of an empirical model is the trapezoidal rule for the integration until the last time point. For larger times, a decay with the physical half-life of the used radionuclide can be assumed (conservative assumption).

The accuracy of the obtained area under the curve depends on the number of measurement points and their distribution over time. An advantage of this model is that no assumptions about the analytical form (or function) of the TAC are included, which could have led to a systematic error. But as shown in Figure 2, there are also disadvantages.

Because of only a few measured time points (possibly not optimally scheduled), numerical integration may lead to errors. The first time point must be chosen properly according to the assumption that must be made for the times between the injection of the radiopharmaceutical and the first measuring time point. A further assumption for the time after the last measurement can also lead to errors.

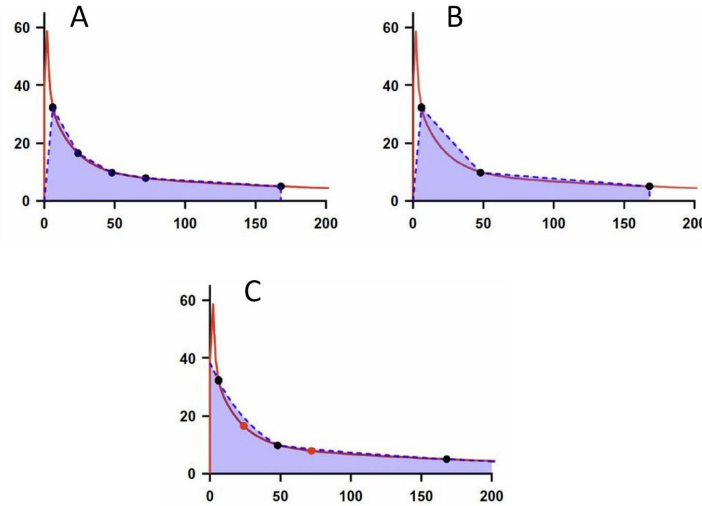


Figure 2. Quantification of the TIAC from a hypothetical TAC. Trapezoidal integration. The X-axis shows the time in hours and the Y-axis indicates the activity in %. The red line describes the true, underlying activity retention. Black dots indicate activity measurements performed at nominal times. Red dots indicate measurements that were omitted from this nominal scheme. A–C trapezoidal integration based on 5 (A), or 3 data points (B, C) where C includes extrapolation before the first and beyond last data point (Sjögreen Gleisner et al. 2022).

Assuming that the relevant biological processes, e.g. the transfer of substances between different compartments, follows the first-order kinetics, it can be shown that TACs can be described by a sum of exponential functions.

$$A(t) = e^{-\lambda_{phys}t} \sum_{i=1}^n A_i e^{-\lambda_i t} \quad (1)$$

$$N = \int_0^{\infty} (e^{-\lambda_{phys}t} \sum_{i=1}^n A_i e^{-\lambda_i t}) dt = \sum_{i=1}^n \frac{A_i}{\lambda_i + \lambda_{phys}} \quad (2)$$

Where $A(t)$ is the activity at time t , $\lambda_{phys} = \frac{\ln 2}{T_{1/2}}$ represents the nuclide's physical decay constant with a half-life of $T_{1/2} = 6.6 d$. A_i and λ_i represent the coefficient and rate constant of exponential term i . Realistic biological processes always have non-linear components. Therefore, the sums of exponential functions are only approximate solutions.

Figure 3 shows that a mono-exponential function is insufficient to adequately represent the true data, and the errors in the TIAC vary greatly depending on which time points are used. Figure 4 shows that the choice of the right points has a decisive role in the calculation of the correct area under the curve (AUC).

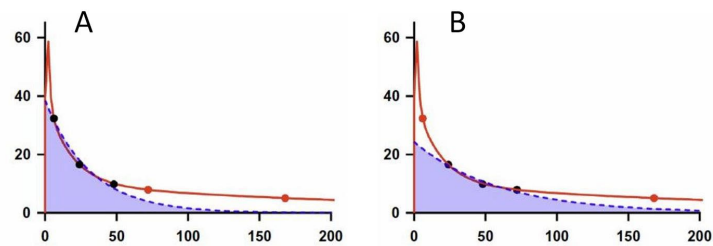


Figure 3. Integration based on mono-exponential curve fitting to 3 data points at different times. The X-axis shows the time in hours and the Y-axis indicates the activity in %. Red dots indicate measurements that were omitted from this nominal scheme (Sjögreen Gleisner et al. 2022).

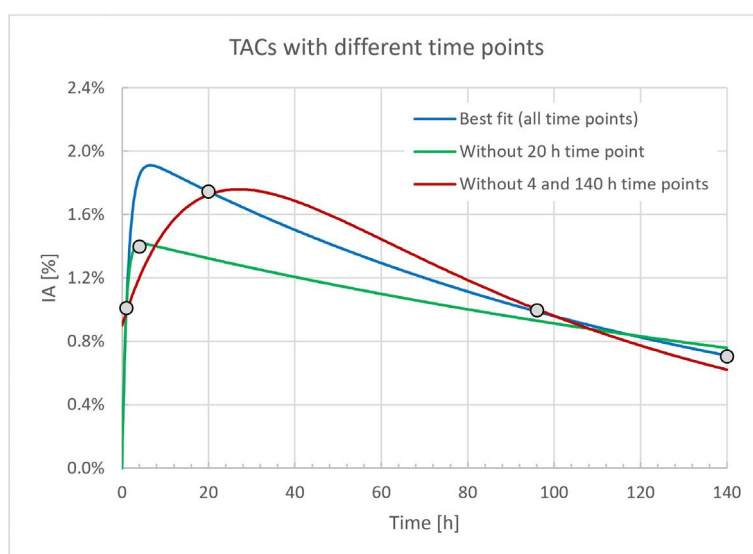


Figure 4. TACs based on the different time points.

Other options are to fit analytical functions, for example, as an appropriate compartmental model function or physiologically-based pharmacokinetic (PBPK) model function, to the data.

Compartment models are based on grouping “similar” spaces and substances to simplify the mathematical description of physiological processes (Jacquez 1996). An organ may be described by a few interconnected compartments or one compartment can also contain several organs. Between these compartments, an exchange of the considered radiopharmaceutical or metabolism occurs (Figure 5). Such compartment model describes only part of the body.

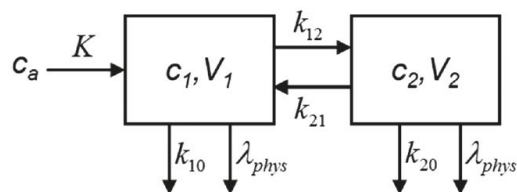


Figure 5. Two-compartment model.

PBPK models reflect (approximately) the actual anatomy and physiology in more detail and encompass the whole body (Figure 6). The PBPK models have considerable knowledge about the anatomy and physiological processes represented by the individual compartments.

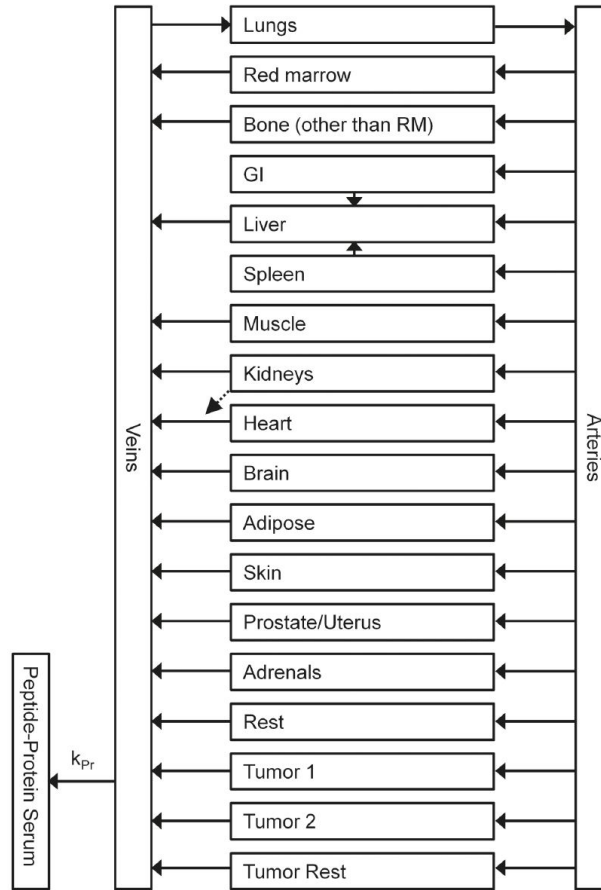


Figure 6. Physiologically based pharmacokinetic whole-body model (Kletting et al. 2016).

The mathematical representation of the movement of the radiopharmaceutical within such a compartment model is described by linear or nonlinear differential equations:

$$\frac{dq_i(t)}{dt} = I(t) - \sum_{j=0, j \neq i}^n (\lambda_{ji} + \lambda_p) q_i(t) + \sum_{j=1, j \neq i}^n \lambda_{ij} q_j(t) \quad (3)$$

where $q_i(t)$ is activity of the radioactive substance in compartment i at the time t ; λ_{ij} is the transfer rate of substance transferred from j to i ; λ_{ji} is the transfer rate from compartment i to j ; λ_{0i} is the loss rate to outside of the system; $I(t)$ is the rate of input from outside the system; and λ_p is the radioactive decay constant.

Mostly investigated models have very few compartments, a consequence of the limited possibilities to measure data from all compartments, the often very small number of measurement data, and the occurring measurement errors.

Due to lack of information and only a few time points in our calculations we used analytical methods and compartmental modelling to find the appropriate fit functions.

In all phases of the TACTIC grand challenge the following software (tools) to find the best fit were used: MATLAB (MATLAB. (2022a). Natick, Massachusetts: The MathWorks Inc.), R, Python, OriginPro

(ver. 2022; OriginLab Corporation, Northampton, MA, USA), SigmaPlot (SigmaPlot Version 12.3 (2013) Systat Software, Inc., San Jose, California), Excel and SAAM II (Barrett et al. 1998).

2.3.1. *SigmaPlot / OriginPro*

SigmaPlot and OriginPro are software packages for data analysis and scientific graphing. Both are very similar and enable extensive curve fitting and statistical analysis. Graphical presentation of data in 2D and 3D is possible. Different type of 2D graph is available, as scatter, contour, radar, dots plots etc.

The data to be fitted can be selected from a worksheet or a graph using a variety of data formats. The fitting equation must have one independent variable. User can specify the behavior of each parameter. A parameter can be localized to have a separate value for each data set, or a parameter can be shared to have the same value for all data sets. Fitting data is easy with the Regression Wizard. The Regression Wizard automatically determines initial parameters, writes a statistical report, saves equation, and adds results to existing graphs or creates a new one. Value of parameters can be restricted in range.

The Regression Wizard accurately fits nearly any equation up to 10 variables and 25 parameters. Weighting is also enabled as well. So far, the user can add his own curve fit equations and add them to the Regression Wizard.

2.3.2. *MATLAB*

MATLAB has multiple built-in functions for data fitting, while many of them are part of specific toolboxes. A quite handy function, regarding the given case, is named *fitnlm()*, which can be found in the Statistical and Machine Learning Toolbox. Abbreviation stands for "fit nonlinear regression model". The function is convenient since adds to results also some statistical metrics like R^2 , adjusted R^2 , AIC, BIC, and RMSE, so no need to evaluate them separately. Missing AICc is not an issue since the simple correction can be used even for non-linear regression models as proposed in (Burnham et al. 2002). Uncomplicated inputs and outputs are also practical for users who are not statistical experts. The minimal input set contains measured data, a regression model (defined as a function), and an initial value vector denoted as *beta0*. The initial guess for a given function can be made by examining the fitted data through 'hand' trial and error or by using bounding conditions (e.g., the regression model value at time zero should not be far from a physiologically meaningful value, and the last data point should be prefitted using all model parameters; the slopes between data points may serve as a basis for exponent parameters). When a regression model is a function of multiple parameters, then their sampling, regression model evaluation, and minimal chi-square (from data) can bring some initial value vectors.

2.3.3. *SAAM II*

SAAM II is user friendly commercial software for compartmental modelling (Barrett PHR, Bell BM, Cobelli C, Golde H, Schumitzky A, Vicini P, Foster DM (1998) SAAM II: simulation, analysis, and modeling software for tracer and pharmacokinetic studies. *Metabolism* 47 (4):484-492. (Barrett et al. 1998) Perazzolo S (2024) SAAM II: A general mathematical modeling rapid prototyping environment. *CPT: Pharmacometrics & Systems Pharmacology*. (Perazzolo 2024) SAAM II can be used in two modules: numerical and compartmental modules, where functions and structured compartmental models can be constructed with variables, i.e. model parameters in the computer program, respectively. These functions and models are then associated with data which can be measured or

adopted from literature and model parameters can be estimated by optimizing the fitting processes. In the following, we introduced the fitting processes using the numerical and the compartmental modules.

2.3.3.1. Curve-fitting with SAAM II numerical module:

The numerical module in SAAM II was designed for a direct curve fitting to time-series data using analytical functions, without the need to define compartmental structures or solve differential equations. This makes it particularly useful for practical applications like mono- or bi-exponential fitting of time-activity curves, such as these time-activity data of [^{177}Lu]Lu-PSMA in blood, kidneys, and tumors. One can follow the following steps to perform the time-curve fitting using numerical module in SAAM II.

Step 1: Starting the numerical module: From the SAAM II main interface, the numerical module can be accessed directly without opening a model diagram. Upon opening, users are prompted to define a dataset, function type, and fitting parameters.

Step 2: Input time-activity data: Time-activity data (e.g., for blood, kidneys, tumor) are manually entered into a file called DATA in the numerical module. Each dataset consists of a series of time points (in minutes or hours) and corresponding activity values (e.g., MBq, %IA or normalized units).

Step 3: Choosing a fitting function: The user can select a predefined analytical function to fit the data. Commonly used functions include: mono-exponential, bi-exponential, tri-exponential or other user-defined expressions. The selected model is displayed, and each parameter is given an initial estimate.

Step 4: Setting initial parameters and constraints: Initial guesses for each parameter are entered manually. The parameter should be defined as “Adjustable” with constraints, such as low and high limits.

Step 5: Performing the Fit: After all inputs are provided, the fitting routine is executed. The numerical module uses nonlinear least-squares optimization to minimize the difference between the model curve and the observed data points.

Step 6: Viewing results: Once fitting is complete, the module provides: Estimated parameter values, residuals and goodness-of-fit metrics. In addition, A graphical plotting of fitted curve vs. data points can be displayed. The results can be putted out for further use, such as, absorbed dose calculations, integration of the area under the curve (similar to TIAC).

The numerical module offers a quick and robust method for fitting TACs when the underlying physiology is well-represented by simple exponential behavior, especially useful for blood clearance, renal kinetics, or tumor washout.

2.3.3.2. Compartmental model fitting with SAAM II compartmental module:

In addition to the numerical model, SAAM II provided “compartmental module”, in which one can construct a compartmental model such as shown in Figure 8. The model is constructed based on the activity at different time points in blood and in organ at risk (in this case, kidneys) and in tumors. The compartment “Rest of Body” (RoB) is added and represents the activity in other organs and tissues which are not explicitly shown in the model. Compartment “bladder” and urine are added according to the human physiological excretion of urine and the transfer rate are taken from ICRP reference values as initial fitting value. The compartments are connected to each other via transfer coefficients, here we call it model parameters. The clinical or experimental data which were read in the DATA can be used for this module. However, these data must be associated with the compartments in the

constructed model with the “sample” indicators. After giving the initial model parameters, one can start to test the fitting processes. The fitting process is a time-consuming work because one should iterate the process by adjusting the model parameters and their value ranges. Generally, one should get a set of model parameters with standard deviations.

Table 1. Transfer coefficients with standard deviation (SD) and coefficient of variation (CV) for the biokinetic model for patient 2 from Phase 0.

Parameter	Value [h^{-1}]	SD	CV (%)
k(1,2)	0.25990	8.34969E-02	32.13
k(1,4)	0.06995	8.54597E-03	12.22
k(1,3)	0.03454	1.52731E-02	44.22
k(2,1)	0.33966	3.85693E-02	11.36
k(3,1)	0.00194	6.98963E-04	35.96
k(4,1)	0.92049	2.06501E-01	22.43
k(5,2)	0.13609	6.39543E-03	4.7

Figure 7 shows an example for fitted curves for patient 2 with the tool SAAM II. For the calculation with data from the patient 2 in Phase 0 the fit function of blood was used as input for the injection in the compartment “blood”. Consequently, 7 parameters, namely k(2,1), k(1,2), k(5,2), k(1,3), k(3,1), k(4,1), and k(1,4), are needed to fit the model to the data (blood, kidneys and tumor). The physical decay constant is $\lambda_{phys} = \frac{\ln 2}{T_{1/2}} = 0.004345 h^{-1}$. The estimated parameters are presented in Table 1.

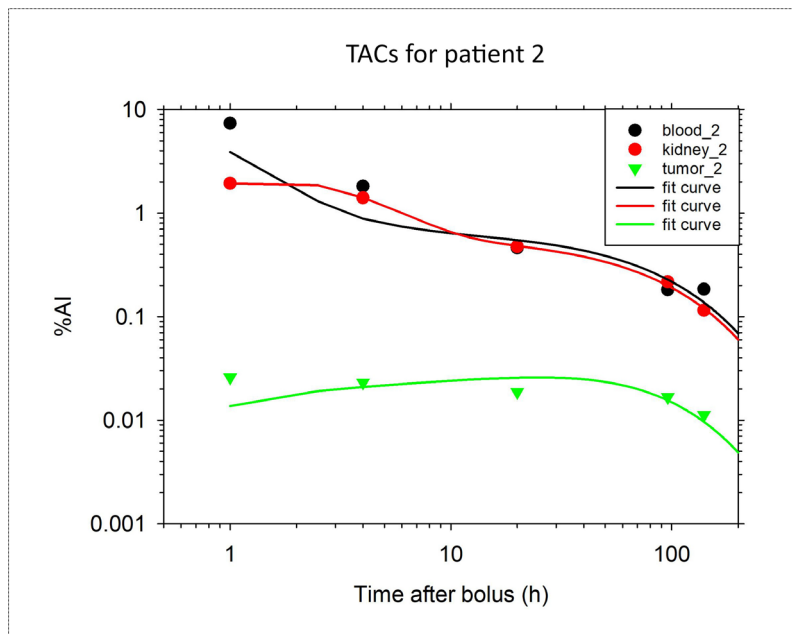


Figure 7. Fitting results for patient 2 (Phase 0) for blood, kidneys, and tumor using compartmental module in SAAM II.

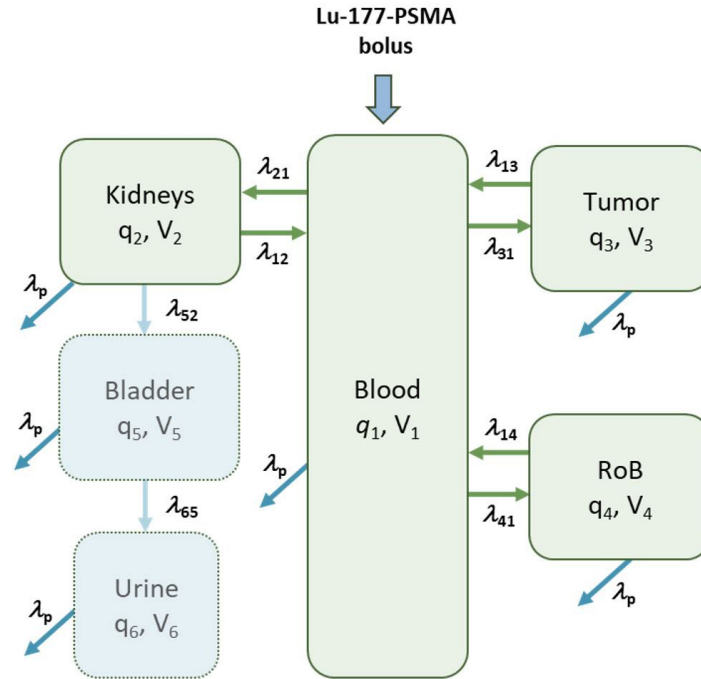


Figure 8. Compartmental model for $[^{177}\text{Lu}]\text{Lu-PSMA-617}$ in therapy.

Results achieved with different tools were compared to find the best parameter sets. Calculations with different tools often provided different results, even if the same function was used. This is since different algorithms, like curve stripping, random search, grid search are used to define optimum starting values. Unlike grid search, for example, random search only tests a certain number of combinations that are selected randomly. As a result, even the same tool can often deliver different values, corresponding to the local minima, when iterated several times. This should be considered when critically analyzing the results.

In all phases for each organ a set of model functions was defined and analyzed, to find the best fit. Determination of the set of functions is an important point in model selection. To get a good model selection output, the functions in the set of model functions should include prior knowledge (Burnham and Anderson 2002), for example, about the physiology of the considered organ. Therefore, characteristics of the functions used in our study based on the prior knowledge of the biokinetic data in kidneys and tumor were the value of $f_i(t = 0) = 0$.

The assumptions made regarding uptake or excretion before the first time point or following the last time point can have a significant impact on the TIAC calculations and should thus be addressed carefully. The choice of the right method of integration can significantly affect the accuracy of the TIAC estimate. Since we are using analytical functions to fit the data, it is obvious to determine the AUC by integration of the function rather than by trapezoid method. It should be considered whether an appearance of an additional component with long half-life after the last time point makes sense. And whether it is useful to determine AUC by integrating the corresponding functions from zero to the last data point and adding the integral from the last data point to infinity using physical decay only (Lassmann et al. 2008).

As no information on the population data was available, no population-based models were considered. In Phase 0 terms like Population-Based Model Selection (PBMS) or Non-Linear Mixed-

Effects Model Selection (NLMEMS) were not used. So, Individual-Based Model Selection (IBMS) was applied to all data in Phase 0.

2.4. Goodness of fit

In Phase 0 simple methods were used to fit the data and decide about the goodness of fit. The visual inspection of the plotted curve and the coefficient of determination R^2 were first only criteria for the test of goodness. Since always several functions are more or less suited to fit given data points, more objective model selection algorithms can be used to reproducibly select the fit function most supported by the data. So, later in Phase 0, different criteria [e.g. F-test (Sachs 1999), the Akaike information criterion (AIC, Akaike 1998), and the Bayesian information criterion (BIC, Schwarz 1978)] were also included.

A visual test serves as an effective diagnostic tool due to its simplicity and ability to identify patterns or irregularities in the data. The estimated parameters of the model should be reasonable. For instance, if they represent physiological values, they must fall within a feasible biological range. Similarly, if the parameters describe transfer rates or biological degradation, they should not have negative values.

2.4.1. Coefficient of determination.

The coefficient of determination, R^2 , represents the fraction of variation in the dependent variable that can be attributed to the independent variables. Its value ranges from 0 to 1, indicating how well the model explains the observed data. The calculation of R^2 follows a specific mathematical formula:

$$R^2 = \frac{SSR}{SST} \quad (4)$$

where SSR is the sum of squares explained by the model and SST is the total variation of the data. A value of R^2 close to 1 corresponds to a good fit. However, an evaluation of R^2 alone may be an inadequate measure of the goodness of fit.

2.4.2. Correlation matrix.

The correlation matrix is a normalized version of the covariance matrix, representing the relationship between pairs of parameters. It is a square, symmetric matrix where the diagonal elements always equal 1. The values within the matrix range from +1, indicating a perfect positive correlation, to -1, denoting a perfect negative correlation. A value of +1 suggests that the two parameters influence the model in a similar manner, whereas -1 indicates an opposite effect. A value of 0 signifies no correlation between the parameters. The correlation matrix is computed using the following formula:

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y} \quad (5)$$

where $\text{Corr}(X, Y)$ is the correlation matrix of parameter X and Y, $\text{Cov}(X, Y)$ is the covariance matrix of parameter X and Y, σ_X is the standard deviation of parameter X, and σ_Y is the standard deviation of parameter Y. High correlation elements tend to widen the confidence intervals. The ideal value of the off-diagonal element of the correlation matrix is close to zero, which shows that the parameters of a function are uncorrelated or independent of each other. Based on the literature, to pass the goodness of fit criteria, most of the elements of the off-diagonal matrix of the fitted parameters of a function should vary from -0.8 to 0.8.

The weighted residuals should be randomly distributed. If they show a systematic trend around the zero line, the model has not captured all information in the data, is biased or incorrect formulated.

To ensure the reliability of the final result, it is essential to assess the uncertainty of the calculated TIAC value. Gear et al. (2018) proposed a guideline addressing uncertainty throughout the entire RPT dosimetry process. In this case, the focus is specifically on uncertainty factors related to TAC fitting. This includes evaluating the underlying noise model used in the data fitting process and the degrees of freedom permitted in the model. Despite often being overlooked and inconsistently reported, uncertainty in TAC fitting is a key quantitative measure that directly reflects the reliability of the TIAC value.

2.4.3. Coefficients of variation (CV) of the fitted parameters.

The coefficient of variation (CV) of the fitted parameters is a statistical metric that quantifies how much a parameter deviates from its average value. It is determined by dividing the standard error by the mean of the fitted parameter. A lower CV indicates a precise estimation of the function parameters, while a higher CV suggests greater uncertainty in their determination. According to literature, a CV below 25 % signifies high precision, whereas values under 50 % are considered acceptable. To meet the goodness-of-fit standards, the CV of fitted parameters should not exceed 50 %.

2.4.4. Statistical methods for model selection.

The F-test, the AIC and BIC criteria represent different theoretical approaches and philosophies.

The F-test (Lomax 2007, p. 10) is only suitable for comparing two nested models. “Nested” means that one model must be a sub-model of the other model. This means that completely independent models cannot be compared. In addition, the restriction to the direct comparison of only two models means that when comparing several models, the result depends on the order in which the two models are compared (Glatting et al. 2007).

AIC allows to quantify the relative Kullback–Leibler distance without knowing the true model by using the maximized loglikelihood function (Akaike 1974; Burnham and Anderson 2002);

$$AIC = -2\ln\hat{L} + 2K \quad (6)$$

where $\hat{L}$ is the likelihood estimator which for normally distributed errors is maximized when the residual sum of squares is minimized and K is the number of estimated parameters included in the model. Our analyses are based on least-squares regression with normally distributed errors and thus the AIC can be computed according to the following formula:

$$AIC = N \cdot \ln\left(\frac{SS}{N}\right) + 2(K + 1) \quad (7)$$

where N is the sample size, K is the number of parameters (+1 for the estimated variance), and SS the sum of squares of the vertical distances of the points from the curve.

Its derivation assumes large sample sizes, which for our case were not available. So, we considered a corrected AIC the AICc, which is the original AIC plus an additional penalty term determined by the sample size and the number of model parameters:

$$AICc = AIC + \frac{2(K+1)(K+2)}{N-K-2} \quad (8)$$

The value of the AICc itself for a given data set has no meaning. The value can only be interpreted by comparing the AICc of a series of models based on the same data set same observations with the same dependent variables. Then, the model with the lowest AICc is the best model among all considered models.

In order to compare models, two measures associated with the AICc can be used: the $\Delta AICc$ and the Akaike weights. The $\Delta AICc_i$ is given by the following formula:

$$\Delta AICc_i = \Delta_i = AICc_i - \min_i(AICc_i) \quad (9)$$

and is the difference between the AICc of model i and of the best model, i.e., the model with the minimal AICc. The greater the $\Delta AICc$ of a model, the less likely is this model the correct model.

Akaike weights w_i represent the ratio of $\Delta AICc$ values for each model relative to the whole set of R candidate models and can be calculated according to the following formula:

$$\text{Akaike weight} = w_i = \frac{\exp(-\Delta_i/2)}{\sum_{r=1}^R \exp(-\Delta_r/2)} \quad (10)$$

The Akaike weights indicate the probability that the model is the best among the whole set of considered models. For instance, a w_i of 0.75 for a model indicates that given the observed data, it has a 75 % chance of being the best among the considered models.

The AICc is a recommended selection criterion, especially in the field of medicine, where reality is very complex, often more than two models are compared, and the sample size is small (Burnham and Anderson 2002). So, in Phase 0, and in later phases as well, we concentrated only on the AIC or AICc.

As can be seen from the Eq. 8 the minimum number of data points for AICc-based model selection is equal to the number of adjustable parameters $K_{max} + 2$. As only five data points for each organ and patient were available, the maximum number of parameters for the calculation of AICc should not exceed 3. In Phase 0, where IBMS was applied to all data and functions with four degrees of freedom provide the best results after visual inspection, AIC only and not AICc could be used for objective evaluation of the goodness of fit.

Similar to AIC, the Bayesian Information Criterion (BIC) is another model selection criterion that considers both model fit and complexity. BIC is based on Bayesian principles and provides a stronger penalty for model complexity compared to AIC for sample sizes greater than 7 (Stoica and Selen 2004). The formula for BIC is as follows:

$$BIC = -2\ln\hat{L} + K \ln(N) \quad (11)$$

In the formula, the terms $\hat{L}$ and K have the same meaning as in AIC and the term $\ln(N)$ represents the logarithm of the sample size N . The $\ln(N)$ term in BIC introduces a stronger penalty for model complexity compared to AIC, as the penalty term scales with the sample size.

BIC overcomes certain limitations associated with the F-test and can accurately identify the correct model when dealing with infinitely large datasets. However, extensive datasets with numerous measured time points are uncommon in nuclear medicine. For smaller datasets, simulations have shown that AIC performs better in comparison. The key distinction between AIC and BIC lies in how they penalize model complexity. AIC applies a penalty proportional to $2K$, while BIC's penalty grows logarithmically with the sample size, imposing a stricter penalty. As a result, BIC is inclined toward selecting simpler models, encouraging a more minimalist approach to model selection. In contrast,

AIC aims to balance model complexity with goodness of fit, whereas BIC enforces a stronger constraint on complexity.

2.5. Error models in the fitting

As no information on uncertainty was provided for the requested data all functions were fitted to the data with different tools or software by a minimum sum of squares algorithm with $1/y$ or $1/y^2$ weighting method. The $1/y^2$ weighting was used to model the errors, as the errors due to pipetting in the blood samples, for example, are larger than the Poisson noise of the radionuclide counting (weighting $1/y$) (Kletting et al. 2009). To investigate the significance of the selected weight, all calculations were performed additionally with weights chosen equal to 1 (no weighting).

Figure 9 shows the influence of the weighting on the fitted curve and correspondingly on the AUC (TIAC). As can be seen from the figure, smaller values are better hit with the weighting $1/y$ or $1/y^2$. Visually, the weighting $1/y^2$ achieves a better fit to the data and correspondingly R^2 values closer to 1. The differences in TIAC values are significant. Nevertheless, in all phases all three cases (1 , $1/y$, $1/y^2$) were examined in the search for the best fit.

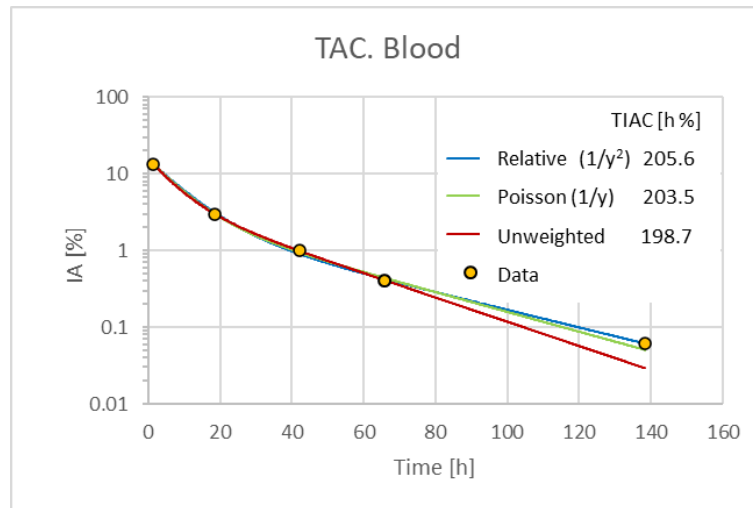


Figure 9. The influence of the weighting on the fitted curve and on TIAC, for blood data for different patients.

If the results do not strongly depend on the chosen weighting, one can trust the result obtained. Later, in Phase 2, as population data was available, typical additive and/or proportional error models inherent in population selection methods could be used (see chapter 4.3).

2.6. Results

2.6.1. Blood

Because biological processes are generally assumed to follow first-order kinetics, which can be described by a sum of exponentials (Strand et al. 1993; Wang et al. 2005), the following functions were used:

$$f_1(t) = Ae^{-\lambda t} \quad (12)$$

$$f_2(t) = A_0 + A_1e^{-\lambda_1 t} \quad (13)$$

$$f_3(t) = A_1e^{-\lambda_1 t} + A_2e^{-\lambda_2 t} \quad (14)$$

$$f_4(t) = A_1 e^{-\lambda_1 t} + (A_0 - A_1) e^{-\lambda_2 t} \quad (15)$$

$$f_5(t) = A [\alpha e^{-\lambda_1 t} + (1 - \alpha) e^{-\lambda_2 t}] e^{-\lambda_{phys} t} \quad (16)$$

$$f_6(t) = A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t} - (A_1 + A_2) e^{-\lambda_3 t} \quad (17)$$

$$f_7(t) = A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t} + (100 - A_1 - A_2) e^{-\lambda_3 t} \quad (18)$$

$\lambda_{phys} = \frac{\ln 2}{T_{1/2}}$ denotes physical decay constant, λ_i are the decay constants of the radiopharmaceuticals (in some cases physical decay constant is included) with values ≥ 0 . The parameter α is the fractional contribution of the corresponding exponentials with values between 0 and 1, and prefactors A account for the total activity.

The duration of the uptake phase in blood may vary widely with different radiopharmaceuticals. The administration time (infusion) in ^{177}Lu -PSMA therapy can take 20-30 min (Zang et al. 2019; Cytawa et al. 2023). But it is also often assumed that uptake of the radiopharmaceutical is instantaneous i.e., the peak uptake occurs immediately following administration.

Figure 10 shows the differences in the TAC for blood for instantaneous (bolus) and a slow administration of ^{177}Lu when the infusion takes 20 min. The simulation was carried out using the biokinetic model of Lutetium published in the ICRP Publication 141 (ICRP, 2019) to demonstrate the fundamental differences in different types of administration of ^{177}Lu . Although it is a simulation of a pure isotope ^{177}Lu , ^{177}Lu -PSMA will behave in a similar way.

A comparison of TACs for blood shows that in the case of 20 min infusion the curve does not reach 100 % of injected activity. The fitting of the data and subsequent integration for these two cases will provide significantly different AUCs.

Thus, the choice of the zero point is decisive and there are several ways to handle the data point at $t = 0$ you do not have. Set the value to zero if you assume that the infusion lasted for some time. Set the value to 100 % if you assume an instantaneous uptake of 100 % of activity. Set the value to < 100 % if you know the value from clinical routine. Extrapolate the observed data back to $t = 0$ or ignore the value at $t = 0$ when fit the data.

As there was no information about the blood activity at time point $t = 0$, the constraints $f_i(t = 0) = 0$ and $f_i(t = 0) = 100$ in the functions f_6 and f_7 , respectively, were considered in Phase 0.

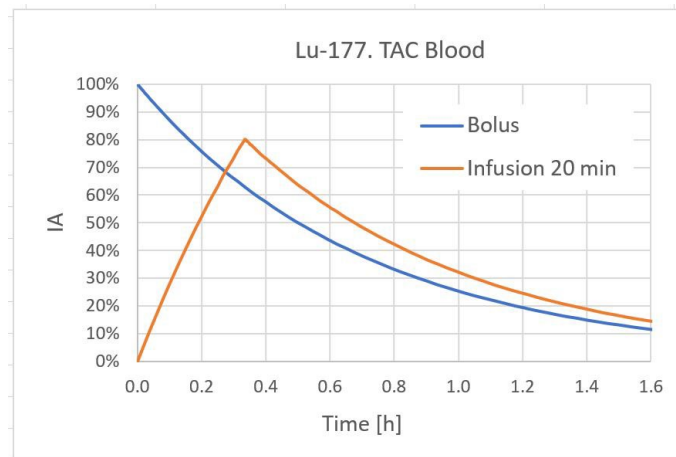


Figure 10. TAC for ^{177}Lu in blood in case of bolus intake or infusion.

The mono-exponential function with two or three parameters f_1 and f_2 , respectively, was used to fit the blood data. The calculated coefficients of determination R^2 for five patients were larger than 0.985. However, the visual inspection did not yield any satisfactory results. Either the value at 4 h or 20 h were clearly outside of the fitted curves. With the exclusion of time points 4 h or 20 h by weighting, it was possible to improve the fitting curve. However, as no information on possible outliers was available, such weighting was not applied and the functions f_1 and f_2 were discarded.

The case in which blood activity reaches 100 % at point $t = 0$ was considered and function f_7 was used to fit the data, but ultimately this case was considered unlikely and therefore discarded. The case in which the point at $t = 0$ would be zero was also discarded as less probable. The point at $t = 0$ was disregarded during fitting in Phase 0.

The visual inspection of the fit with functions f_3 , f_4 and f_5 showed good results. The calculated R^2 for all three functions was close to 1.

The integration of the fitted function from 0 to infinity was applied as considerable deviations are reported if the last time-point is well above zero, i.e., the conservative assumption of only physical decay afterward is often not justified (Siegel et al. 1999; Konijnenberg 2011).

The fitting was carried out with the function *fitnlm()* in MATLAB, and validation was performed with software packages OriginPro and SAAM II. For a given function, different software packages provided different parameters. These parameters were used as initial values in MATLAB to start iteration. In this way, with cross comparison between different tools, parameters in MATLAB were found that provided the best AIC.

The settings of the *fitnlm()* function were as follows: constant error model, maximum iteration count of 200, termination tolerance for the objective function value of 1E-8, termination tolerance for the parameters of 1E-8, and the Levenberg-Marquardt nonlinear least squares algorithm.

The best AIC was achieved with the bi-exponential function f_3 with four adjustable parameters, shown in Table 2. All parameters are given with standard errors (SE). The value at $t = 0$ was not considered, only five given data points were used when fitting the data. The results for time-integrated activities are shown in Table 3. In Phase 0 we didn't divide the values by 100, so that the values given are percentages.

A Monte Carlo simulation would be necessary to analyze the uncertainty of the calculated TIACs. This would make it possible to determine the statistical distribution of the TIACs with confidence intervals and statistical variables such as mean values or standard deviations. Since the Monte Carlo simulation is very time-consuming, it was decided to calculate the standard errors in time-integrated activity by means of uncertainty propagation. For the function $y = f(x_1, x_2, \dots, x_N)$ the squared uncertainty is calculated according to the following formula:

$$u^2(y) = \sum_{i=1}^N \left(\frac{\partial f}{\partial x_i} \right)^2 u^2(x_i) + 2 \sum_{i=1}^{N-1} \sum_{j=i+1}^N \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} u(x_i, x_j) \quad (19)$$

where $u^2(x_i)$ denotes the variance of the parameter x_i and $u(x_i, x_j)$ is the covariance of the parameters x_i and x_j . The standard error (SE) of the parameter or the function is then equal to $u(x_i)$ or $u(y)$, respectively. In our case the covariance term was equal to 0, as no correlation between the parameters was assumed.

Figure 11 shows the fits for blood for the first 5 patients from Phase 0, performed with the function f_3 and parameters given in Table 2.

Table 2. Fit parameters for the function f_3 for blood data with standard errors (SE).

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5
A_1 [%]	15.3426	11.8422	16.8378	16.3713	23.182
SE	0.1353	0.2187	0.1567	0.3883	0.3021
λ_1 [h ⁻¹]	0.4022	0.5545	0.3341	0.2515	0.3799
SE	0.0093	0.0217	0.0061	0.0063	0.0110
A_2 [%]	2.0015	0.548	2.5405	7.9524	4.6714
SE	0.1764	0.0847	0.1715	0.4121	0.3942
λ_2 [h ⁻¹]	0.0324	0.0094	0.0294	0.0423	0.0373
SE	0.0040	0.0026	0.0029	0.0023	0.0040
R^2	0.9999922	0.9999176	0.9999917	0.9999981	0.9999943
AIC	-21.3016	-14.7480	-19.1850	-23.0206	-17.7718

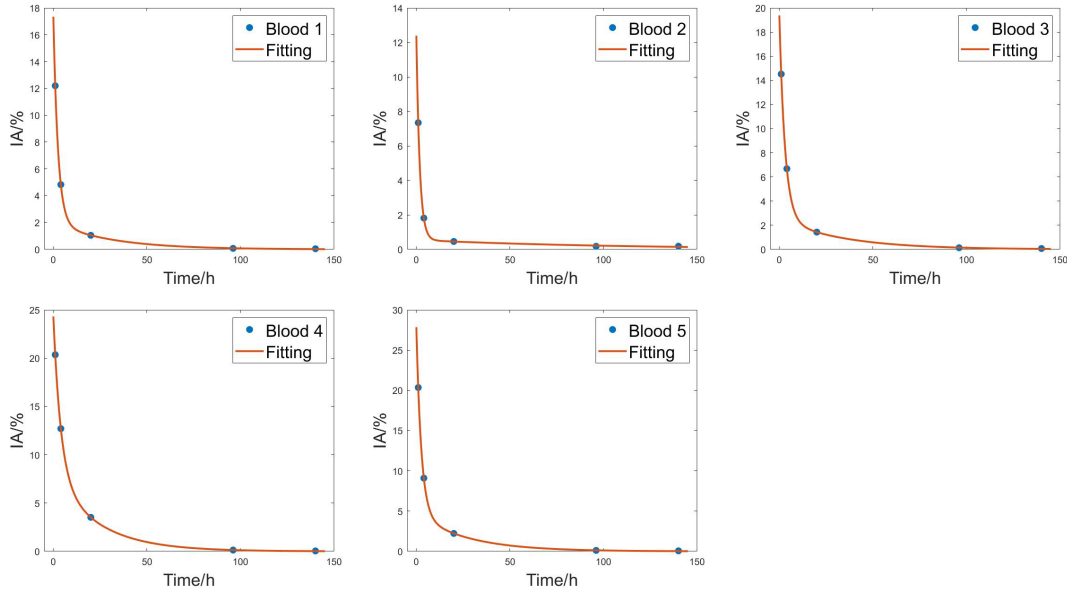


Figure 11. Fitted blood data from five separate patients

Table 3. Calculated TIACs [%] for blood for patients 1-5 with standard errors (SE) [%].

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5
TIAC	99.92	79.65	136.81	253.09	186.26
SE	9.42	18.49	10.38	14.30	17.20

2.6.2. Kidneys

The following sums of exponential functions with 3, 4, and 5 parameters were used to fit the biokinetics of [^{177}Lu]Lu-PSMA in kidneys:

$$f_3(t) = A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t} \quad (20)$$

$$f_{4d}(t) = A_1 \frac{e^{-\lambda_{phys} t} \{ (1-\alpha) e^{-\lambda_1 t} - \alpha e^{-\lambda_2 t} - (1-2\alpha) e^{-\lambda_{bc} t} \}}{\left\{ \left(\frac{1-\alpha}{\lambda_1 + \lambda_{phys}} \right) - \left(\frac{\alpha}{\lambda_2 + \lambda_{phys}} \right) - \left(\frac{1-2\alpha}{\lambda_{bc} + \lambda_{phys}} \right) \right\}} \quad (21)$$

$$f_6(t) = A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t} - (A_1 + A_2) e^{-\lambda_3 t} \quad (22)$$

$$f_8(t) = A_1 (e^{-\lambda_1 t} - e^{-\lambda_2 t}) \quad (23)$$

Function f_3 only delivered good results when weighting was applied to individual time points. However, as with blood data, as no information on possible outliers was available, such weighting was not applied and the functions f_3 was discarded.

It was decided to use prior knowledge of the biokinetic data in kidneys and apply the value of $f_i(t = 0) = 0$.

The function f_{4d} with four adjustable parameters was used, as it was selected as the function most supported by the kidneys data (Hardiansyah et al. 2023). λ_{phys} is the physical decay constant of ^{177}Lu

($\lambda_{phys} = \frac{\ln 2}{T_{1/2}} = 0.004345 \text{ h}^{-1}$), λ_{bc} is the rate of blood circulation in humans of 1 min ($\lambda_{bc} = \frac{\ln 2}{1 \text{ min}} = 41.59 \text{ h}^{-1}$), λ_i are the biological decay constants of the radiopharmaceutical with values ≥ 0 , the α value is the fractional contribution of the corresponding exponentials with value between 0 and 1, and prefactors A accounts for the total activity.

Radioactive decay λ_{phys} should be included explicitly into the equations to avoid possible artifacts for example due to incorrect weighting factors (Williams et al. 1995; Glatting and Reske 1999).

All fit parameters should be constrained to values greater than 0. However, if the physical decay constant λ_{phys} is included in the decay constants, as in functions f_3 , f_6 or f_8 , then the decay constants should be constrained to values greater than λ_{phys} .

To avoid these possible sources of errors in Phases 1 and 2 the physical decay constant should be included explicitly into equations for all used functions.

Figure 12 shows an example for TACs, performed for kidney data from patient 2 with 3 different tools, R (A), MATLAB (B) and OriginPro (C). The calculations were performed with function f_6 with 5 adjustable parameters with calculated values for R^2 : 0.99968, 0.99985, 0.99986; AIC: -44.10, -48.87, -49.16, and TIAC: 55.91, 58.91, 62.44.

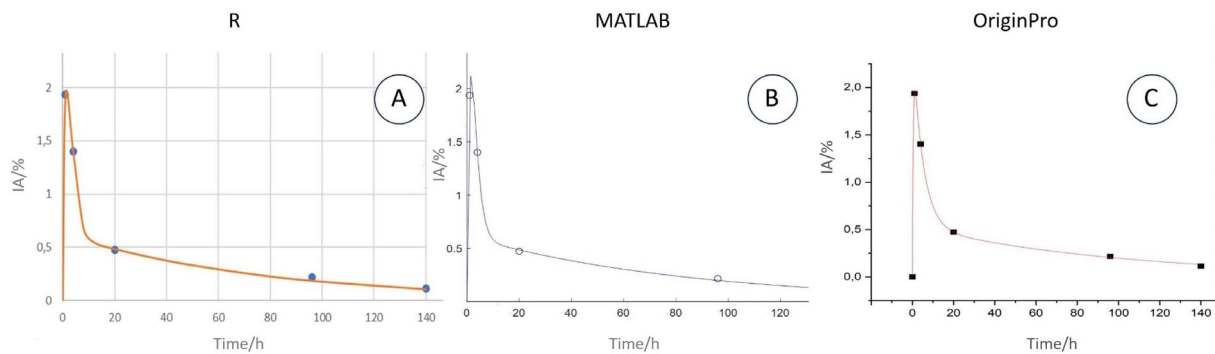


Figure 12. TACs for kidneys, for patient 2, performed with function f_6 in R (A), MATLAB (B), and OriginPro (C).

Compared with other functions, after visual inspection and comparison of R^2 and AIC, the function f_{4d} provided the best results. The calculated values for R^2 were in the range between 0.9974 and 0.9999.

The best fits were achieved with parameterization, shown in Table 4.

Figure 13 shows the fits for kidneys for the first 5 patients from Phase 0, performed with the function f_{4d} and the results for time-integrated activities (in percent) with standard errors are given in Table 5.

Table 4. Fit parameters for the function f_{4d} for kidneys data with standard errors (SE).

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5
A_1 [%]	76.3804	59.9486	172.0005	129.9536	61.6449
SE	5.5878	2.9063	4.3291	6.2584	1.1266
α	0.6583	0.9927	0.6469	0.7513	0.4321
SE	0.0325	0.0036	0.0237	0.0996	0.0089
λ_1 [h^{-1}]	0.1090	0.0067	0.0124	0.0771	1.6206
SE	0.0650	0.0008	0.0006	0.0838	0.0655
λ_2 [h^{-1}]	0.0152	0.6903	2.1402	0.0128	1.0230
SE	0.0028	0.0190	0.3240	0.0029	0.1217
λ_{bc} [h^{-1}]	41.59	0.65	0.5	41.59	0.0056
R^2	0.9974	0.9999	0.9999	0.9989	0.9999
AIC	-17.1230	-27.0107	-20.2282	-16.4581	-37.3203

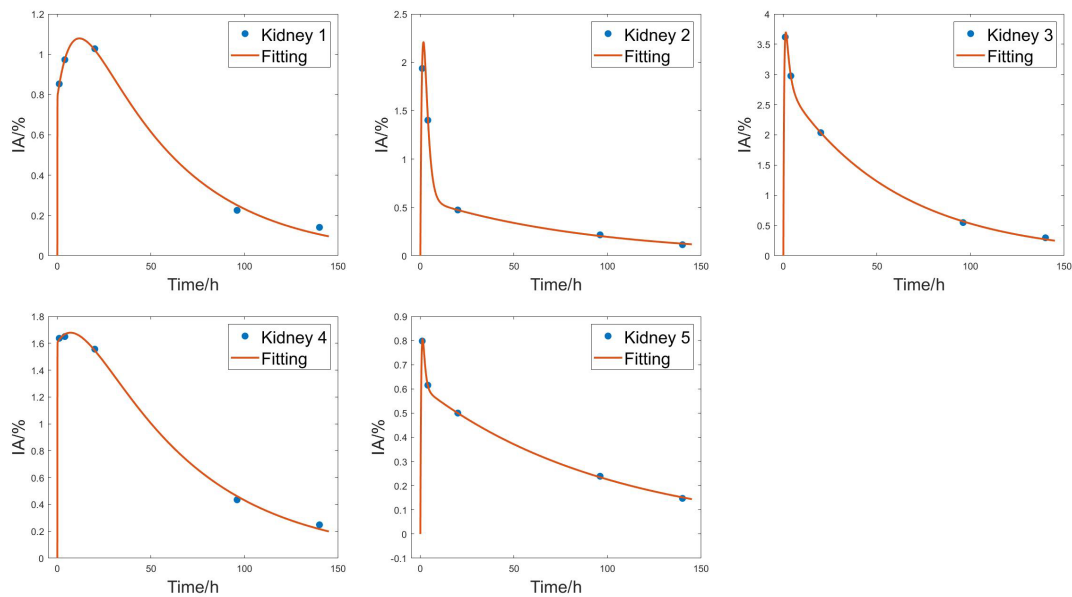


Figure 13. Fitted kidneys data from five separate patients.

Table 5. Calculated TIACs [%] for kidneys for patients 1-5 with standard errors (SE) [%].

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5
TIAC	76.38	59.95	172	129.95	61.64
SE	5.5878	2.9063	4.3291	6.2584	1.1266

2.6.3. Tumor

As with the kidneys, the following exponential functions were used in the tumor as well:

$$f_3(t) = A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t} \quad (24)$$

$$f_{4d}(t) = A_1 \frac{e^{-\lambda_{phys} t} \{ (1-\alpha) e^{-\lambda_1 t} - \alpha e^{-\lambda_2 t} - (1-2\alpha) e^{-\lambda_{bc} t} \}}{\left\{ \left(\frac{1-\alpha}{\lambda_1 + \lambda_{phys}} \right) - \left(\frac{\alpha}{\lambda_2 + \lambda_{phys}} \right) - \left(\frac{1-2\alpha}{\lambda_{bc} + \lambda_{phys}} \right) \right\}} \quad (25)$$

$$f_6(t) = A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t} - (A_1 + A_2) e^{-\lambda_3 t} \quad (26)$$

$$f_8(t) = A_1 (e^{-\lambda_1 t} - e^{-\lambda_2 t}) \quad (27)$$

Of the three given organs, the tumor data was the most difficult data to fit, as in some cases the values did not follow a logical progression. As can be seen in Figure 14, for patients 1 and 4, for example, the values at time point 4 h are lower than at the time points 1 h and 20 h. Assuming that the value at the time point 4 h is an outlier, the point was “excluded” from the fitting procedure by weighting and good results could be achieved with bi-exponential functions with three (patients 1, 3, 4 and 5) and four parameters (patient 2). However, since it was not known whether the value was an outlier, it could not be treated as one, which made the fitting procedure more difficult. Finding a well-parameterized fit function for the tumor data of all five patients was a real challenge.

A visual inspection of the tested functions f_3 , f_6 and f_8 with calculated R^2 did not provide reasonable results. As with the kidneys, the decision was made to use the f_{4d} function here as well. Compared with other functions, f_{4d} was identified as the function that provided the best fits. The visual inspection of the fitted function delivered acceptable results. The calculated values for R^2 were in the range between 0.9219 and 1.

Table 6. Fit parameters for the function f_{4d} for tumor data with standard errors (SE)

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5
A_1 [%]	4.9347	5.0201	2.2960	0.8047	2.6271
SE	1.1796	1.6758	0.2107	0.3116	0.0109
α	0.5783	0.7479	0.5632	0.5003	0.6229
SE	0.1558	0.2886	0.0903	0.4245	0.0012
λ_1 [h ⁻¹]	0.0557	0	0.0317	0.0344	0.1740
SE	0.1396	0	0.0281	0.0056	0.0033
λ_2 [h ⁻¹]	0.0102	1.4363	0.0105	0.0343	0.0033
SE	0.0168	0.4416	0.0084	0.0056	0.0000
λ_{bc} [h ⁻¹]	41.59	5.0201	41.59	41.59	41.59
R^2	0.9219	0.9393	0.9903	0.9588	1.0000
AIC	-34.9710	-44.5919	-56.1834	-45.7898	-87.2252

With the parameterization for the function f_{4d} , shown in Table 6, the best results, with the smallest AIC, could be achieved.

Figure 14 shows the fits for tumor for the first 5 patients from Phase 0, performed with the function f_{4d} . Table 7 shows the results for time-integrated activities for each patient (in percent) and the corresponding standard errors.

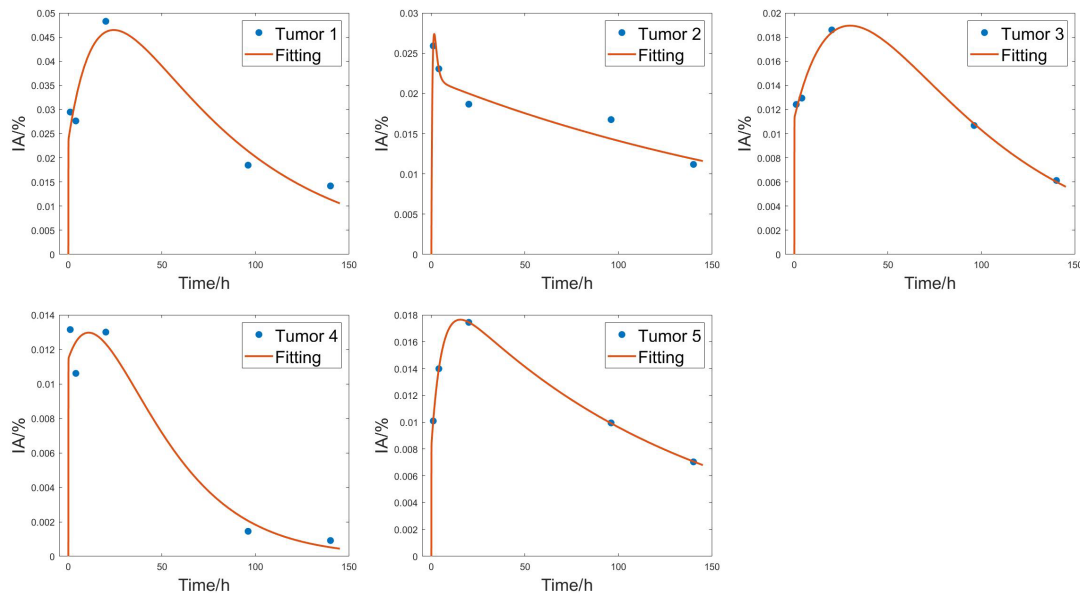


Figure 14. Fitted tumor data from five separate patients.

Table 7. Calculated TIACs [%] for tumor for patients 1-5 with standard errors (SE) [%].

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5
TIAC	4.93	5.02	2.296	0.805	2.63
SE	1.1796	1.6758	0.2107	0.3116	0.0109

2.7. Discussions

There was a very good agreement among the EURAOS team members on the following data points: 4-, 20-, 96- and 140-hours post administration. A serious discussion happened on the first datapoint. Larger part of the team vouched for the 1-hour post application since relayed on biokinetic modeling experience which tries to capture early uptake phase if possible. The time points 1, 24, 48, 72 and 168 hours also apply as five-time point SPECT protocol reference standard in some publications (Peters et al. 2023). Few members propagated later time for the first data point, i.e. between 2 to 3 hours rather than 1 hour as the studies have shown that imaging at 2-3 hours post-injection provides more accurate and reliable dosimetry results for both kidneys and tumors (Peters et al. 2023). It was finally considered to request the first time point at 1 hour rather than between 2 and 3 hours. In retrospect, however, when looking at the graphs (Figures 11-14), one can see that there would be no advantages in choosing the earlier time point differently.

A large set of functions would be useful when searching for the appropriate results. However, the testing of a large number of functions requires a lot of time, which is not always possible. In our cases we used only seven different functions for the blood data and only four for kidneys and tumor data.

Despite the small number of tested functions, it was managed to achieve good results. According to minimum root-mean-square error (RMSE) and maximum deviation of the results from the ground truth TIAC values of the synthetic data, the second place for kidneys with $RMSE = 4.01$ and the first place in final scoring with $RMSE = 51.66$ was achieved. One of the reasons for this was that results from the literature were used where similar cases had already been investigated. So, for example, the function f_{4d} with four adjustable parameters was used for kidneys and tumor data, as it was selected as the function most supported by the kidneys data from the literature (Hardiansyah et al. 2023).

When selecting possible candidates, it is important to use parameters in the function that have a certain physical meaning inherent in the data under consideration. It is often incorrectly assumed that the fit function must have fewer parameters than there are measurement data. The fit function may have more parameters, but fewer parameters can be fitted than there are measurement data. Thus, several parameters need to be fixed based on a priori knowledge. For example, physical decay is a known parameter of the fit function that can be factored out. Another example would be the parameter λ_{bc} , the rate of blood circulation in humans, that was used in the function f_{4d} for the kidneys and tumor data. This rate of blood circulation in humans of 1 min is a known parameter and should also be fixed when fitting the data. Such parameters should be inserted in the function explicitly, as they can potentially improve the fitting results.

It was found that all the software used delivered similar results. What distinguishes the software, however, is the difficulty to find the “best” result. So, for example, to find the best parameters in OriginPro, it’s important to start the iterations with a good set of initial values. Ultimately, our approach was to start to find the best parameters according to the AIC values with all tools in parallel, to share the parameters between the tools as initial values and to try it again. In this way, the best results were finally found.

When searching for the best function or the best set of parameters not only AIC values but also BIC values were compared. BIC uses a different concept, which means that other parameter sets are preferred in relation to the BIC values. However, since the decision was made to use AICs and Akaike weighting factors, as the best selecting criteria, the BICs were ultimately ignored.

Finding a fitting function for the blood data was straightforward. The data were well distributed and without outliers. The choice of an early time point at 1 h and the last time point that was available at 140 h allowed the early rapid clearance and long-term retention of the exponential function to be covered well. The only real concern was the lack of information at time 0. It was decided to ignore the time point 0. So, in Phase 0 five given data point were considered when fitting the function and the integration of the fitted function from 0 to infinity was applied. The conservative assumption of only physical decay after last time point was not taken into consideration as it was not justified.

The example in Figure 12 shows how sensitive the calculation of TIACs is. TACs performed with 3 different tools, R, MATLAB and OriginPro, look almost identical. The calculated R^2 are similar and the AICs differ only slightly. But the differences in calculated TIACs (more than 10 %) are significant.

For kidneys data after comparison of four possible candidates the function f_{4d} with the parameter λ_{bc} provided the best results. In Phase 0, we were not yet aware that parameter λ_{bc} should be fixed to the value 41.59 h^{-1} , so we varied the parameter, as this allowed us to achieve better results. This is the reason why the values in Table 4 for patients 2, 3 and 5 differ from the value 41.59 h^{-1} . Physically, however, it makes no sense; we should have fixed the parameter and possibly introduce an additional parameter that we could then vary. Nevertheless, with parameters given in Table 6 best

results could be achieved and in Phase 1 more attention should be paid to the fact that the parameters have a physical meaning.

Tumor data looked similar to kidneys data but were more difficult to fit. It was and still is not entirely clear how to deal with such data like values at time point 4 h of patient 1 and 4. Are these outliers, as they are smaller than the values at time points 1 h and 20 h, and should be excluded from the fitting process? Or should this fact be ignored and all given values should be used when fitting the data? A comparison of the results without and with excluded data with the “true values” could give an answer. Detecting outliers is an essential process across various domains, including healthcare. It involves pinpointing data points that deviate significantly from the rest of the dataset. Outliers may signal errors in measurement, unusual patterns, or even fraudulent activities. In the medical field, this technique is valuable for identifying patients with atypical symptoms, uncovering irregularities in billing, or detecting possible equipment failures. There are numerous mathematical methods in descriptive statistics for outlier detection. One commonly used method is the Z-score method (Zhang et al. 2014). This method calculates the difference between each data point and the mean of the dataset, and divides it by the standard deviation. If a data point has a Z-score greater than a certain threshold, it is considered an outlier. Another method is Interquartile Range (IQR) Method (McClave and Sincich 2016). This method identifies outliers based on the range between the 25th and 75th percentile. Any value outside 1.5 times the IQR is flagged as an outlier.

However, to detect the outliers more information about data is needed. That wasn't the case for us. So, in our case a decision was made never exclude data, neither explicitly nor by weighting, so that in all phases all given data were always included in the fitting process. As for kidneys, we have also for tumor for patient 2 nonsensical values for parameters α , λ_1 and λ_{bc} (Table 6). However, as the best results could be achieved with these values, we ignored the physical sense at this point and used the values as they were.

2.8. Conclusions

The warm-up Phase 0 was a good exercise to prepare for the following tasks in the next phases. The tasks set were discussed in detail. The choice of data points was crucial, as it meant that all phases of the curve could be well covered. The data points were successfully fitted. Good results were achieved by comparing and using different fitting methods and software packages.

The submitted results were scored and ranked by the Challenge Team according to minimum root-mean-square error (RMSE) and maximum deviation of the results from the ground truth TIAC values of the synthetic data. With the fitted kidneys data, the 2nd place could be reached. Together with blood and tumor results finally the best result of all participants could be achieved. However, as the RMSE value was quite high and consequently the deviation from the true value was large, many questions remained unanswered at the end of Phase 0. Some of these are: What value should be assumed at time $t = 0$? What is the shape of the curve after the last data point?" What are the best selection criteria for the fit function?

These questions should then be attempted to be answered in Phases 1 and 2.

3. Grand Challenge Phase 1

3.1. Introduction and motivation

Accurate estimation of radiation dose delivered to organs and tumors is fundamental to the success of radiopharmaceutical therapy. Central to this process is the reliable modeling of TACs, which describe the kinetics of radiopharmaceutical uptake and clearance in tissues over time. These TACs are subsequently integrated to estimate the TIAC, a critical parameter for personalized dosimetry. In contrast to Phase 0, where participants could freely select time points and familiarize themselves with data handling and modeling, Phase 1 introduced a more structured and competitive setting. All participants received synthetic TAC data for blood, kidneys, and tumors from five individual patients, with predefined sampling schedules. This phase challenged participants to optimize multiple components of the dosimetry workflow: model selection, e.g., multi-term exponential functions, fitting procedures, such as least-squares or weighted regression, and data handling methods, all aimed at achieving high accuracy in organ-level TIAC estimation. Importantly, participants were required to perform curve fitting without use of population-based modeling or information sharing across patients, emphasizing the effectiveness and robustness of exclusively individualized modeling approaches. This constraint made Phase 1 a realistic simulation of clinical scenarios where patient-specific data are sparse, and no prior population knowledge is available. Phase 1 also served as a critical benchmark for evaluating the potential benefit, or trade-offs, of incorporating population-level information in subsequent phase. By establishing a standardized evaluation framework centered on individual fitting accuracy, Phase 1 not only revealed the methodological diversity among participants but also underscored the need for harmonization and best-practice guidelines in personalized dosimetry. It highlighted how performance variability can arise from differences in model assumptions, fitting strategies, and the treatment of uncertainty, even when all participants are working with identical datasets. Therefore, Phase 1 played an introductory role in the challenge's overarching goal: to promote consensus, transparency, and reproducibility in TAC-based dosimetry methods for radiopharmaceutical therapy.

3.2. Data

In Phase 1, data for further five patients for kidneys, tumor and blood and for five time points was provided by the organizers.

Figure 15 shows the percent (%) activity data for kidneys, tumor, and blood from five separate patients at the provided time points. The data points are connected by a dashed line to better distinguish individual patient data from each other. As in Phase 0, no information about data uncertainty was provided.

As in Phase 0, the data show a similar progression. There is a decrease for kidneys, blood and some patients for tumor and an increase and subsequent decrease for other patients for tumor. The only difference to the data progression from Phase 0 is that in Phase 0 there was also an increase for kidneys for some patients, followed by a decrease. So unlike Phase 0, in Phase 1 a mono-exponential function with 2-3 degree of freedom might be sufficient to fit the kidneys data.

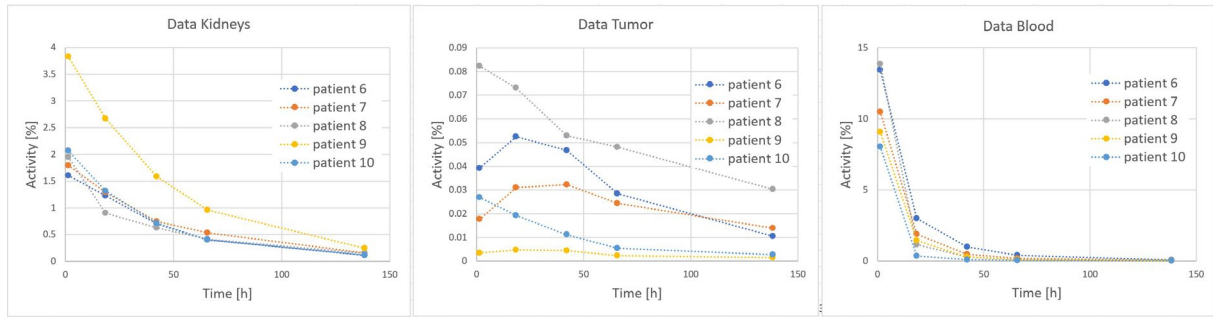


Figure 15. Percent (%) activity data for kidneys, tumor and blood from five separate patients.

3.3. Consider value at time point 0 for blood data

The functions in the set of model functions should include prior knowledge, for example, about the physiology of the considered organ. In Phase 0 we considered the prior knowledge of the biokinetic data in kidneys and tumor were the value at the time point zero was set to zero ($f_i(t = 0) = 0$). However, for blood data the value at time point zero was ignored and only five given time points were considered when fitting the data. The function f_3 was considered for blood data, as it was the best choice in Phase 0. With function f_3 the best results were achieved in Phase 0 and since the blood data in Phase 1 were similar to the data in Phase 0, it made sense to use the function to fit data again. However, in Phase 1 the question was raised as to what exactly happens at time point 0 in blood data and that it might make sense not to ignore the value at time point zero.

Assuming that all of the activity is immediately released into the blood when $[^{177}\text{Lu}]\text{Lu-PSMA}$ is injected, 100 % of the administered activity could be assumed at time point 0 (the term “volume of distribution” was not yet familiar at that time).

In Phase 0 the case in which blood activity reaches 100 % at point $t = 0$ was considered unlikely and therefore discarded. Many publications address blood clearance and have resulted in different values (Kabasakal et al. 2017; Khawar et al. 2019). However, studying more existing publications on this or similar topics resulted in the realization that possibly 100 % of the administered activity reaches the blood (Abuqbeitah et al. 2018; Santos-Cuevas et al. 2018; Carpanese et al. 2020). So, in Phase 1 it was decided to consider this case for blood clearance again hoping that this will improve the results by calculating the TIACs.

3.4. Results

3.4.1. Blood

Based on the experience gained from Phase 0, the number of functions was reduced and the following functions were selected for a possible fit:

$$f_3(t) = [A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t}] e^{-\lambda_{phys} t} \quad (28)$$

$$f_9(t) = [A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t} + (B - A_1 - A_2) e^{-\lambda_3 t}] e^{-\lambda_{phys} t} \quad (29)$$

λ_{phys} denotes physical decay constant, λ_i describe the biological clearance rates of the radiopharmaceutical with values ≥ 0 , prefactors A account for the total activity and factor B accounts for the total activity at time point 0.

As mentioned in Phase 0 to avoid possible sources of errors in Phases 1 and 2 the physical decay constant is included explicitly into equations for all used functions.

For such high values at time point 0 it would be impossible to find a good fit to the blood data with the function f_3 . The function with at least five parameters were needed. So, function f_9 was considered and by selecting parameter B accordingly, the cases with initial activity in blood of 100 % or similar could be realized. Also, as mentioned in Chapter 2 a gradual administration of the radiopharmaceutical in the [^{177}Lu]Lu-PSMA therapy with slow infusion, that takes 20-30 min, could be expected.

To find a fit function with appropriate parameters, under consideration of all these possible conditions, proved difficult to achieve. The comparative calculations have shown that an instantaneous uptake of 100 % of activity could be a good assumption for high activity values at time point 0 in blood. So, function f_9 with parameter B = 100 was used, to find the parameter set for the best fit.

The best fits were achieved with parameterization, shown in Table 8 and Figure 16 show the fits for blood for 5 patients, performed with the function f_9 with weighting $1/y^2$. The calculated values for R^2 were equal to 1 for all patients. Unfortunately, the AIC values were not noted in the rush. Thus, Table 8 does not contain AICs. The results for time-integrated activities are shown in Table 9.

Table 8. Fit parameters for the function f_9 for blood data with standard errors (SE).

	Patient 6	Patient 7	Patient 8	Patient 9	Patient 10
A_1 [%]	6.9731	6.4949	3.7635	5.5833	1.4442
SE	0.0016	0.0028	0.0016	0.0030	0.0040
λ_1 [h^{-1}]	0.0510	0.0735	0.0620	0.0743	0.0814
SE	3.04E-5	4.67E-5	3.55E-5	5.78E-5	2.05E-4
A_2 [%]	0.6820	0.4702	0.0625	0.1939	0.0804
SE	0.0030	0.0013	4.01E-4	0.0014	5.15E-4
λ_2 [h^{-1}]	0.0134	0.0127	8.01E-5	0.0135	0.0058
SE	2.98E-5	2.82E-5	5.37E-5	7.50E-5	6.35E-5
λ_3 [h^{-1}]	1.9130	2.2075	1.5876	2.2737	1.9180
SE	1.65E-4	4.92E-4	8.75E-5	5.67E-4	3.37E-4
R^2	1	1	1	1	1

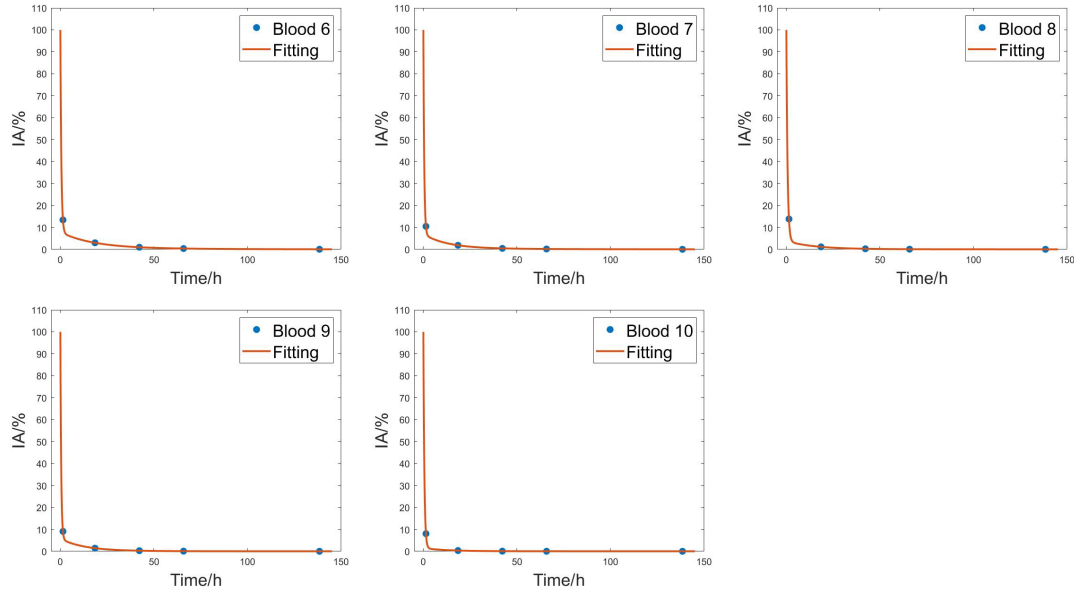


Figure 16. Fitted blood data from five separate patients.

Table 9. Calculated TIACs for blood for patients 6-10 with standard errors (SE).

	Patient 6	Patient 7	Patient 8	Patient 9	Patient 10
TIAC	2.1315	1.5307	1.3128	1.2323	0.7598
SE	0.00196	0.00108	0.00198	0.00111	0.07925

3.4.2. Kidneys

Based on the experience gained from Phase 0, the following exponential functions with 4-5 degrees of freedom were selected for a possible fit:

$$f_3(t) = [A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t}] e^{-\lambda_{phys} t} \quad (30)$$

$$f_{4d}(t) = A_1 \frac{e^{-\lambda_{phys} t} \{ (1-\alpha) e^{-\lambda_1 t} - \alpha e^{-\lambda_2 t} - (1-2\alpha) e^{-\lambda_{bc} t} \}}{\left\{ \left(\frac{1-\alpha}{\lambda_1 + \lambda_{phys}} \right) - \left(\frac{\alpha}{\lambda_2 + \lambda_{phys}} \right) - \left(\frac{1-2\alpha}{\lambda_{bc} + \lambda_{phys}} \right) \right\}} \quad (31)$$

$$f_6(t) = [A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t} - (A_1 + A_2) e^{-\lambda_3 t}] e^{-\lambda_{phys} t} \quad (32)$$

$$f_{10}(t) = [A_1 t e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t}] e^{-\lambda_{phys} t} \quad (33)$$

The functions f_3 and f_6 were modified by inserting of physical decay constant λ_{phys} explicitly in the equation (compare equations 14, 17 with 28, 32) to avoid possible artefacts. The function f_8 (Eq. 23) was discarded in Phase 1 as it had proved unsuitable in Phase 0.

The function f_{10} was additionally tested as possible candidate which is defined in analogy to the case of degenerate eigenvalues for a damped oscillator (note the additional factor t in the first term of the equation).

The function f_{4d} with four adjustable parameters was used, as it was selected as the function most supported by the kidneys data (Hardiansyah et al. 2023) and because it delivered the good results for kidneys in Phase 0.

The fitting was carried out with the function $fitnlm()$ in MATLAB with the same settings as in Phase 0. The validation was performed with software packages OriginPro and SAAM II. For a given function, different software packages provided different parameters. These parameters were used as initial values in MATLAB to start iteration. In this way, with cross comparison between different tools, parameters in MATLAB were found that provided the best AIC.

At $t = 0$, the activity in kidneys was set to zero as a boundary condition. The best AIC was achieved with the function f_{4d} when the weight was set to $1/y$. The calculated values for R^2 were in the range between 0.9945 and 0.9994.

The calculation of TIACs was also performed with the function f_3 . The calculated values were as follows: patient 6: 0.8115, patient 7: 0.9898, patient 8: 0.8056, patient 9: 1.8462 and patient 10: 0.8537.

Table 10 and Figure 17 show the parameterization and the best fits for kidneys for 5 patients, performed with the function f_{4d} , and in Table 11 are shown the results for time-integrated activities with corresponding standard errors.

Table 10. Fit parameters for the function f_{4d} for kidneys data with standard errors (SE).

	Patient 6	Patient 7	Patient 8	Patient 9	Patient 10
A_1 [%]	82.1313	96.5468	78.8370	187.2720	86.6927
SE	2.7571	3.9506	1.6511	3.1352	4.4274
α	0.5762	0.5266	0.0232	0.5000	0.4954
SE	0.5542	0.7161	0.0317	0.5872	0.8719
λ_1 [h^{-1}]	1.7664	19.2429	1.5609	82.8154	21.1190
SE	1.4100	0.0173	0.1674	0.0018	0.0210
λ_2 [h^{-1}]	0.0169	0.0143	0.0110	0.0165	0.0195
SE	0.0016	0.0011	0.0006	0.0005	0.0017
λ_{bc} [h^{-1}]	41.59	41.59	1.45	41.59	41.59
R^2	0.9963	0.9945	0.9994	0.9991	0.9918
AIC	-12.4436	-9.8389	-21.7594	-13.3942	-7.16701

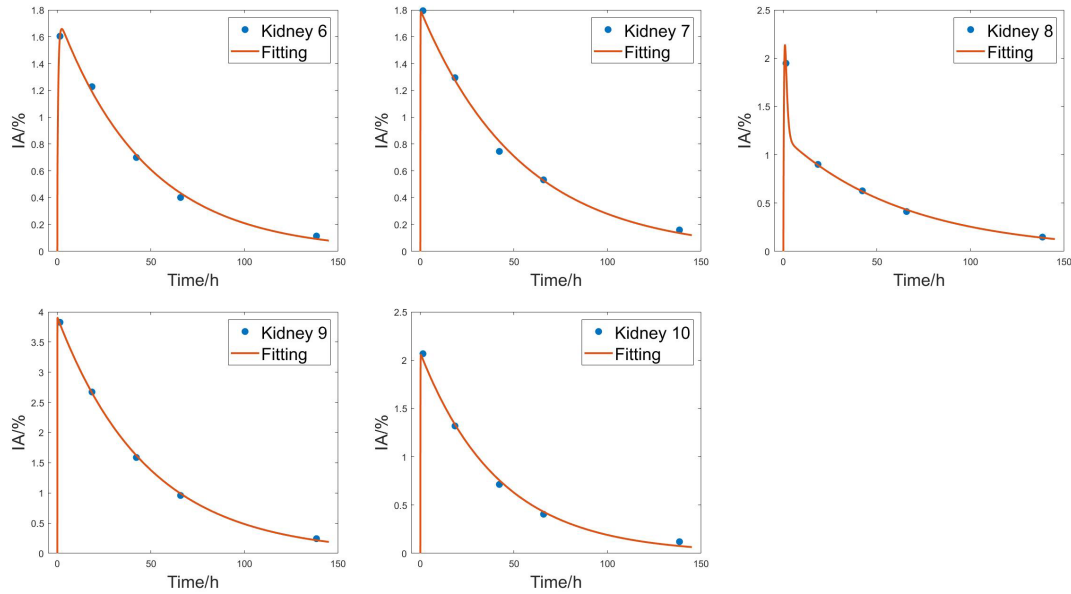


Figure 17. Fitted blood data from five separate patients.

Table 11. Calculated TIACs for kidneys for patients 6-10 with standard errors (SE).

	Patient 6	Patient 7	Patient 8	Patient 9	Patient 10
TIAC	0.8213	0.9655	0.7884	1.8727	0.8667
SE	0.0276	0.0395	0.0165	0.0314	0.0443

3.4.3. Tumor

Because good results were achieved in Phase 0, the following exponential functions with 4 degrees of freedom were selected for the tumor data:

$$f_{4d}(t) = A_1 \frac{e^{-\lambda_{phys}t} \{ (1-\alpha)e^{-\lambda_1 t} - \alpha e^{-\lambda_2 t} - (1-2\alpha)e^{-\lambda_{bc} t} \}}{\left\{ \left(\frac{1-\alpha}{\lambda_1 + \lambda_{phys}} \right) - \left(\frac{\alpha}{\lambda_2 + \lambda_{phys}} \right) - \left(\frac{1-2\alpha}{\lambda_{bc} + \lambda_{phys}} \right) \right\}} \quad (34)$$

The fitting was carried out with the function *fitnlm()* in MATLAB, and also here validation was performed with software packages Origin, SAAM II, and IDL.

At $t = 0$, the activity in tumor was set to zero as a boundary condition, and the best AIC was achieved when the weight was set to $1/y$. The calculated values for R^2 were in the range between 0.9010 and 0.9898.

Table 12 and Figure 18 show the parameterization and the best fits for tumor data for 5 patients, performed with the function f_{4d} . The results for time-integrated activities are shown in Table 13.

Table 12. Fit parameters for the function f_{4d} for tumor data with standard errors (SE).

	Patient 6	Patient 7	Patient 8	Patient 9	Patient 10
A_1 [%]	4.9439	4.8210	10.6379	0.5296	1.2905
SE	0.4422	0.5309	0.0023	0.1572	0.1224
α	0.6283	0.5975	0.5000	0.6663	0.5000
SE	0.0483	0.0228	0.0002	0.1412	0.0003
λ_1 [h^{-1}]	0.0747	0.0704	0.0034	0.1426	38.3004
SE	0.0595	0.0361	0.0335	0.0586	0.0003
λ_2 [h^{-1}]	0.0112	0.0045	0.0034	0.0066	0.0169
SE	0.0035	0.0019	0.0322	0.0061	0.0027
λ_{bc} [h^{-1}]	41.589	41.589	41.589	41.589	41.589
R^2	0.9898	0.9854	0.9707	0.9010	0.9702
AIC	-41.1632	-48.1117	-35.1718	-55.9622	-45.9189

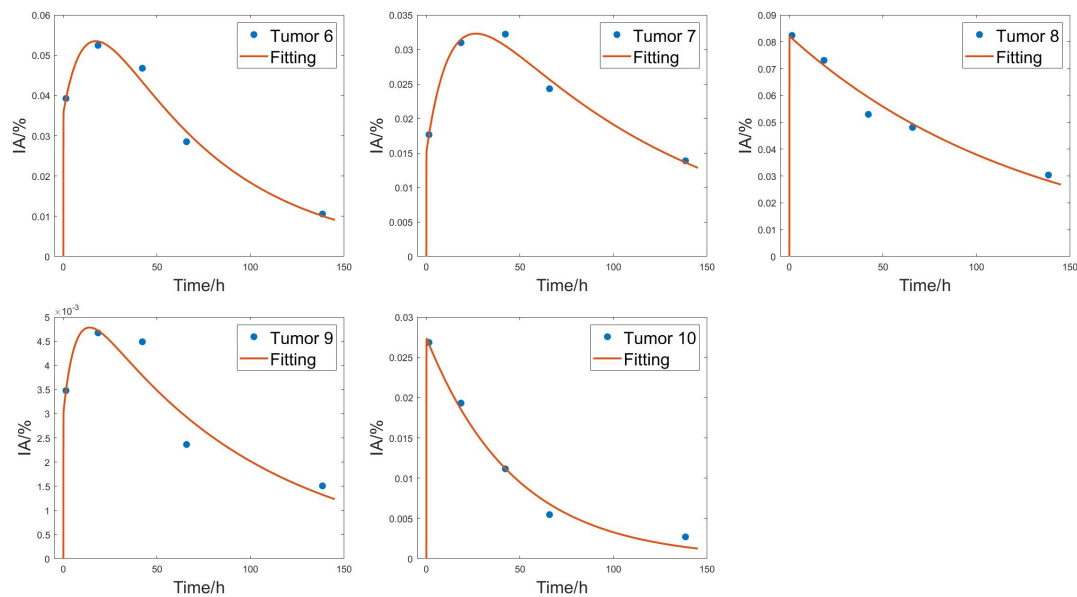


Figure 18. Fitted tumor data from five separate patients.

Table 13. Calculated TIACs for tumor for patients 6-10 with standard errors

	Patient 6	Patient 7	Patient 8	Patient 9	Patient 10
TIAC	0.0494	0.0482	0.1064	0.0053	0.0129
SE	0.0044	0.0054	2.3E-5	0.0016	0.0012

3.5. Discussion

In Phase 1, data were available for 10 patients (5 from Phase 0 and 5 from Phase 1), so it would made sense to use the population-based models when fitting the data. However, we should apply the population-based approaches only in Phase 2 in order to take advantage of the opportunity to experiment with individual-based models in Phase 1. This mainly concerned the blood data, where an attempt was made to apply different constraints in the fitting functions than in Phase 0.

According to minimum root-mean-square error (RMSE) and maximum deviation of the results from the ground truth TIAC values of the synthetic data, the 4th place for kidneys with very good RMSE of 0.046 could be achieved. However, in final scoring the results were unsatisfactory (not in top 5). Since the data for tumor were treated the same as in Phase 0, it was obvious to believe that the poor results were due to the results for blood data.

As in Phase 0 finding a fitting function for the blood data in Phase 1 for patients 6-10 was straightforward. The data were well distributed and without outliers. No comparison with other functions was carried out for blood data. So, the AICs for different models were not compared and not noted (Table 8). Only one point under discussion was what is the activity at time zero. Numerous publications on the subject of blood activity at time zero led us to assume that 100 % of the activity can actually be present in the blood at time zero. So, in contrast to Phase 0, it was decided to assume 100 % blood activity at time point zero in Phase 1. In retrospect, this proved to be the wrong decision.

When [¹⁷⁷Lu]Lu-PSMA is administered, the volume of distribution V_d , the theoretical volume that would be necessary to contain the total amount of an administered drug at the same concentration that it is observed in the blood plasma, is apparently quite big.

The volume of distribution can be written as:

$$V_d = \frac{A}{C} \quad (35)$$

where A is amount of drug in the body (dose, shortly after administration) and C is plasma concentration.

The higher the volume of distribution, the larger is the amount of activity distributed outside of the blood plasma. The main causes of a large V_d are distribution in peripheral tissues (muscle, fatty tissue, bone, etc.) and plasma protein binding. For patients 1-5 in Phase 0 at time zero a maximum value of 30 % were achieved and due to similar behavior of blood data in Phase 1, similar values at time zero should be expected.

As mentioned before the administration time (infusion) in [¹⁷⁷Lu]Lu-PSMA therapy can take more than 20 min and after injection [¹⁷⁷Lu]Lu-PSMA rapidly leaves the blood and localizes in target tissues. These make it unsuitable to assume that 100 % of the injected activity is in the blood plasma at time zero. Obviously if the volume of distribution is not measured and no information about activity at time point zero is available the time point zero should be ignored when fitting the blood data.

As in Phase 0 for kidneys data after comparison of four possible candidates the function f_{4d} with the parameter λ_{bc} provided the best results. The decay parameters λ for all patients were constrained to positive values, parameter α , the fractional contribution of the corresponding exponentials, was in range between 0 and 1 and the parameter λ_{bc} , with exception of patient 8, was fixed to the value 41.59 h⁻¹ (Table 10). For patient 8 was difficult to find an appropriate fit function with $\lambda_{bc} = 41.59$ h⁻¹. It was therefore decided to vary the parameter λ_{bc} in order to achieve the best possible results.

The data of patient 4 and especially patient 1 from Phase 0 showed a clear uptake. So, it was obvious when fitting the data to use a sum of at least two exponential functions, to cover uptake and retention ranges. In Phase 1, however, the kidneys data showed no uptake pattern. If you ignore the fact that the biokinetic data in kidneys and tumor were 0 at time point 0 ($f_i(t = 0) = 0$), you could choose the simpler function than f_{4d} , mono-exponential function or function like f_3 that contains few parameters, to find the best fit. All the functions f_3 , f_{4d} and f_6 delivered good results, but the smallest AICc in f_{4d} was ultimately decisive to choose f_{4d} as the best candidate.

Good results were also achieved when fitting the kidneys data with the function f_3 . Here the time point 0 was ignored and only 5 given time point were used. The calculation of TIACs was performed by integrating of the function from first time point to the infinity and the area between the time zero and the first time point was calculated according to the trapezoidal rule. The calculated TIACs for patient 6 (0.8115), patient 7 (0.9898), patient 8 (0.8056), patient 9 (1.8462) and patient 10 (0.8537) were similar to the results performed with the function f_{4d} . Nevertheless, the calculated AICs were smaller for the fit function f_{4d} . Thus, this fit function with parameterization shown in Table 10 was identified as the best one.

As in Phase 0 tumor data looked similar to kidneys data but were more difficult to fit. Nevertheless, good results were achieved with the same function as for the kidneys and all constraints for the parameters (Table 12) were fulfilled.

3.6. Conclusion

The Phase 1 was another good exercise to learn to find the optimal results when fitting the patient data. The data points for all organs were successfully fitted and if one had not experimented with the blood data, one of the best results among all participants could be achieved.

The decision to set the amount of activity at time point 0 to 100 % in blood data turned out to be wrong and strongly distorted the calculated TIACs. The term “volume of distribution” is decisive when thinking about amount of distributed activity in plasma immediately after injection. If the volume of distribution is not measured or known the better decision is to ignore the value at time point zero as to set it to 100 % of injected activity or other value when fitting the blood data.

Compared to Phase 0 the received data in Phase 1 were more harmonious and well distributed. The kidneys data showed no uptake. For this reason, on the one hand there were more functions as possible candidates, to find the best fit, on the other hand it was easier to build the fits as it was easier to find better initial values when starting the iteration.

Also, the tumor data showed more logical pattern over the time scale in Phase 1 then in Phase 0. The was no behavior of tumor data in Phase 1 like in patient 1 or 4 from Phase 0, where data at some time point looked like outliers. This made it easier to find the appropriate parameters in the fit function.

As in Phase 0 different tools were also used in Phase 1 and good results were achieved by comparing and using different fitting methods and software packages. Also, here the submitted results were scored and ranked by the Challenge Team according to minimum root-mean-square error (RMSE) and maximum deviation of the results from the ground truth TIAC values of the synthetic data. With the fitted kidneys data, the 4nd place could be reached, but because of the wrong decisions when fitting the blood data, not the best final result could be achieved.

However, the Phase 1 gave answers to many questions from Phase 0 and was another good exercise to gain knowledge about fitting techniques. But the population methods should be investigated and used in the last phase of the grand challenge, in Phase 2.

4. Grand Challenge Phase 2

4.1. Introduction and motivation

In Phase 1, elements of population-based modeling were implicitly introduced through introducing the multiple synthetic patients. However, the TAC fitting itself was performed in isolation, with an emphasis on modeling each individual patient independently. In contrast, Phase 2 explicitly extended the TAC fitting framework into a more advanced and clinically realistic context by fully incorporating population-based modeling. This phase explored the central question: To what extent can population priors improve individualized dosimetry, particularly in cases with limited, irregularly sampled, and patient data with large uncertainties? Participants in Phase 2 were provided with a larger, synthetic cohort comprising TAC datasets from 15 new patients, designed to reflect a wide range of physiological variability. This dataset enabled all participants to adopt advanced modeling strategies that extended beyond traditional individual curve fitting and participants should improve the curve fitting for the 5 patients received in Phase 1. Several references were introduced as well for guidance. Approaches such as mixed-effects models encouraged participants to leverage inter-patient information, such as identifying shared kinetics, common parameter distributions, or applying population-based modeling techniques. Phase 2 simulated a data-rich setting similar to well-established clinical protocols or multi-center studies, where historical or concurrent patient data are available to support modeling efforts. This paradigm shift required participants to consider their strategy once again: rather than optimizing a model for each patient in isolation, they need to balance fidelity to individual data with recognition of global patterns and population-level consistencies. In this context, Phase 2 evaluated not only the algorithmic performance of each approach, but also the participant's ability to generalize and transfer knowledge across datasets. A key motivation was to determine whether population-based approaches could mitigate the effects of sparse sampling and improve the precision of dosimetry, particularly when individual measurements are limited or uncertain. Moreover, Phase 2 highlighted the diversity in implementation of population-based modeling, ranging from simple empirical priors to fully probabilistic, data-driven models. These different methodologies revealed important trade-offs between model complexity, interpretability, and predictive accuracy. Therefore, Phase 2 served as a test bench for forward-oriented dosimetry workflows, demonstrating how prior knowledge and artificial intelligence can be integrated into clinical decision-making. Phase 2 challenged participants to transition from conventional data fitting toward model-based learning, marking a critical step forward in the evolution of personalized dosimetry in nuclear medicine. In conclusion, building on the individualized modeling methods established in Phase 1, Phase 2 provided a comparative perspective on the strengths and limitations of individual versus population-based approaches, finally deepening our understanding of variability, uncertainty, and optimization in TAC modeling and consequently the patient dosimetry in radiopharmaceutical therapy.

4.2. Data

Predefined time-activity data for 15 more patients were provided from organizers to improve the fitting results in Phase 1. For patient No. 11 to 20, the five time points were 1.6 h, 19 h, 42.5 h, 66.3 h, and 163.8 h. For patient No. 21 to 25, the fifth time points were instead at 186.5 h (for patient No. 21 to 23), 210.9 h (for patient No. 24), and 235.7 h (for patient No. 25).

Table 14 shows the time points for all 25 patients received from the organizers in Phases 0, 1, and 2.

Table 14. Time points received from organizers in Phases 0-2.

	Phase 0	Phase 1	Phase 2			
Time [h]	Patients 1-5	Patients 6-10	Patients 11-20	Patients 21-23	Patient 24	Patient 25
1	•					
1.4		•				
1.6			•	•	•	•
4	•					
18.5		•				
19			•	•	•	•
20	•					
42.2		•				
42.5			•	•	•	•
65.8		•				
66.3			•	•	•	•
96	•					
138.4		•				
140	•					
163.8			•			
186.5				•		
210.9					•	
235.7						•

4.3. Discussion on individual and population model

In nuclear medicine, the measurement of biokinetics is often only carried out at a few time points. In this challenge, too, only 5 time points were available per patient and organ. This means that functions with a limited number of parameters are available as possible candidates for fitting the data. In addition, fitting the 5 time points with functions with 4 degrees of freedom by using the individual-based model selection (IBMS), the calculation of AICc is not possible.

Therefore, in Phase 2 instead of using the data of only a single patient, the inclusion of the data of patients of Phase 1 with the same disease treated with the same radiopharmaceutical might be a better approach to determine an optimal fit function. Such a population-based model selection (PBMS) increases the ratio of number of observed data used as input to the number of estimated parameters and thus reduces the uncertainty in the model selection. Moreover, it allows to use an expanded model set, as functions with a higher number of parameters and calculation of AICc become possible.

In Phase 2, we therefore used PBMS approach to improve the calculation of TIACs using biokinetic data of a population of Phase 1 instead of a single patient only. For this purpose, a set of mathematical models or functions was defined, a population-based fit was performed and the function most supported by the data was selected using the Akaike weights method.

There are various methods for population-based model selection, i.e. the PBMS can be used in different ways.

4.3.1. PBMS (Population-based model selection)

There are infinite possible curves or functions that “fit” a pharmacokinetic dataset to a certain degree. To estimate the values of the adjustable parameters an objective function, a measure of the fit to the data, depending on the data, the parameters of the error model, the fitting function and (if required and available) Bayesian information, is minimized.

The objective function given in Eq. 36 is an extended version of the least square maximum likelihood (Muzic and Christian 2006; Kletting et al. 2013). It is derived from the product of the likelihood of the observed data (time activity data) and the prior distribution, i.e., *a priori* knowledge of parameter values which may be available from previous experiments or publications. This *a priori* knowledge about the parameter values of the fit function is integrated as additional sums of squares as follows:

$$-2 \ln(P) = \sum_{i=1}^N \left\{ \ln(2\pi\theta_i^2) + \ln(v) + \frac{(y_i - f(t_i))^2}{\sigma_i^2} \right\} + \sum_{j=1}^K \left\{ \ln(2\pi\omega_j^2) + \frac{(p_j - \bar{p}_j)^2}{\omega_j^2} \right\} \quad (36)$$

where P is the likelihood, i denotes the data points, N is the total number of data points, j denotes the parameters, K is the total number of parameters, y_i and $f(t_i)$ are the measured data and the function values at time t_i , θ_i^2 is the variance, v is the scaling factor, σ_i^2 is the scaled variance of data point i , p_j is the current value of parameter j , $\bar{p}_j$ is the average population value, and ω_j^2 is the pertaining variance.

The Bayes terms (right sum in Eq. 36) correspond to additional measurements of the parameters with the given uncertainty. As it is possible to construct an infinite number of fit functions to be used in Eq. 36, it is mandatory to find the “fit function most supported by the data”. Usually, several fit functions are tried out and often the best suited fit based on experience is selected.

4.3.2. SP-PBMS (Shared-parameter population-based model selection)

In shared-parameter population-based model selection a shared parameter approach is used (Hardiansyah et al. 2021). The shared parameters are assumed to be the same for all patients and are estimated for all data in the patient population together. The other parameters were individually estimated from the data.

To investigate if the data of the patients could be described by shared parameters, the population fitting with PBMS approach could be performed to estimate the parameters of functions with shared parameter estimation.

4.3.3. NLMEMS (Non-Linear Mixed-Effects model selection)

Another population-based approach is the non-linear mixed-effects modelling. Nonlinear mixed effects models have been the primary analysis framework for population-based pharmacometric modeling and can be used for accurately determining time-integrated activities.

Parameters in the NLME model consist of the fixed and random effects. Fixed-effect parameters are model elements that take on a particular, or scalar, value. No matter how it is employed in a model, a fixed-effect parameter takes on a single, fixed value. This value is updated during the estimation process until convergence is obtained. The final fixed-effect parameter values of the model are those that minimize the objective function described above or meet other stopping criteria of the estimation method.

For population models, random effects may describe the magnitude of the variability in the values of parameters between subjects (BSV), i.e., inter-individual variability.

$$P_i = P_{pop} + \eta_i \quad (37)$$

$$\eta \in N(0, \omega^2) \quad (38)$$

$$y = f(t, \theta, N(0, \omega^2)) \quad (39)$$

where P_i is the estimated parameter i in an exponential function, P_{pop} represents the typical value for the population (fixed effect), and η_i the random effect. η is a random variable following a Gaussian distribution with mean zero and variance ω^2 . y is a function of time t that describes the model. It depends on vector of fixed effects θ and the normal distributed random effects η .

If the random effects additionally describe the magnitude of the variability between occasions within a subject (BOV), i.e., inter-occasion variability, for the k^{th} occasion Eq. 37-39 become as follows:

$$P_{ik} = (P_{pop} + \kappa_k) + \eta_i \quad (40)$$

$$\kappa \in N(0, \pi^2) \quad (41)$$

$$y = f(t, \theta, \eta \in N(0, \omega^2), \kappa \in N(0, \pi^2)) \quad (42)$$

where κ is random effects following a Gaussian distribution with mean zero and variance π^2 .

Finally random effects may describe the magnitude of unexplained differences between the predicted and observed values of the dependent variable. It is called the residual unknown variability (RUV), often referred to in the literature as intra-individual variability.

$$\epsilon \in N(0, \sigma^2) \quad (43)$$

$$y = f(t, \theta, \eta \in N(0, \omega^2), \kappa \in N(0, \pi^2)) + \epsilon \quad (44)$$

where ϵ is additive random variable describing observation randomness not explained by the model. It follows a Gaussian distribution with mean zero and variance σ^2 that may or may not be a function of time.

The following equations describe typically used error models: the additive or constant (Eq. 45), proportional (Eq. 46), combined (additive + proportional) (Eq. 47), and exponential (Eq. 48):

$$y(t) = f(t) + \epsilon(t) \quad (45)$$

$$y(t) = f(t) \times (1 + \epsilon(t)) \quad (46)$$

$$y(t) = \epsilon_1(t) + f(t) \times (1 + \epsilon_2(t)) \quad (47)$$

$$y(t) = f(t) \times e^{\epsilon(t)} \quad (48)$$

For the model itself, the exponential is probably the most commonly used. These models describe variability of the parameter which is log-normally distributed. A model for the parameters P_i from Eq. 40 becomes:

$$P_i = P_{pop} \times \exp(\eta_i) \quad (49)$$

P_i then is log-normally distributed if the η_i is assumed to be normally distributed.

4.3.4. NLME-PBMS (Population-Based Model Selection within the framework of the Non-Linear Mixed-Effects)

Population-based model selection can be used within the framework of the non-linear mixed-effects modeling. The fixed effects could describe the mean values of the estimated parameters in the population, while random effects describe the inter-patient variability of the estimated parameters between subjects in the population.

The mean values of the estimated parameters in the population could also be estimated by means of shared parameter approach. So, different methods could be combined here to achieve the best results.

4.4. Results

4.4.1. Blood

As it turned out that the assumption we did in Phase 1 that the amount of injected activity at timepoint 0 is 100 % is not correct, the same exponential function with 4 parameters as in Phase 0 was used for fitting of TAC in blood data in Phase 2:

$$f_3(t) = [A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t}] e^{-\lambda_{phys} t} \quad (50)$$

Table 15 and Figure 19 show the parameterization and the best fits for blood data for 5 patients from Phase 1, performed with the function f_3 , respectively.

Table 15. Fit parameters for the function f_3 for blood data, NLME AICc = -43.575.

	Patient 6	Patient 7	Patient 8	Patient 9	Patient 10
A_1 [%]	0.3839	0.3916	0.3182	0.2709	0.1473
λ_1 [h^{-1}]	0.0090	0.0109	0.0118	0.0151	0.0105
A_2 [%]	10.9515	9.9712	12.0579	9.2236	9.1959
λ_2 [h^{-1}]	0.0580	0.0851	0.1218	0.0943	0.1811

The fittings were performed using the non-linear mixed-effects model function *nlmefitsa()* in MATLAB. The optimization was performed using the *fminsearch* algorithm with a maximum of 100,000 iterations and 100,000 objective function evaluations. The convergence criterion was based on default settings of the algorithm, and the error model used was exponential. The method for approximating the loglikelihood was "importance sampling".

The function was applied to 20 patients of Phase 1 and 2 together to find the best parameter set for function f_3 for patients 6-10 of Phase 1. The fixed effect values with standard errors for blood data are shown in Table 16. The random effects for blood data and the covariance matrix for the random effects are shown in Tables 17-18, respectively. Here the model defines random effects additively. Thus, the individual parameters are computed as $X_i = X_{fixed} + \eta_i$, where X_i denotes the fit parameters from Table 15, X_{fixed} denotes the fixed effect parameters from Table 16, and η_i are the random effect values from Table 17.

The results for time-integrated activities with corresponding standard errors are shown in Table 19. To calculate the standard errors for TIACs, we used the variances (diagonal elements) from the covariance matrix (Table 18) and applied them to the data from Table 15 while conducting the uncertainty propagation.

Since the TIACs for patients 6-10 from Phase 1 were calculated with completely different boundary conditions (100 % at time point 0), no comparison with the Phase 1 results can be made here.

Table 16. Fixed effect and standard errors (SE) for blood data.

	Fixed effect	SE
A_1 [%]	0.3349	0.0405
λ_1 [h^{-1}]	0.0099	0.0012
A_2 [%]	8.7835	0.7580
λ_2 [h^{-1}]	0.1178	0.0103

Table 17. Random effects for blood data.

	Patient 6	Patient 7	Patient 8	Patient 9	Patient 10
A_1 [%]	0.048974	0.056708	-0.01676	-0.06401	-0.18764
λ_1 [h^{-1}]	-0.00094	0.000991	0.001887	0.005235	0.000644
A_2 [%]	2.168046	1.187696	3.274423	0.440154	0.412414
λ_2 [h^{-1}]	-0.05977	-0.03273	0.004008	-0.02353	0.06324

Table 18. Covariance matrix for the random effects.

	A_1	λ_1	A_2	λ_2
A_1 [%]	0.016657	0	0	0
λ_1 [h^{-1}]	0	1.91E-05	0	0
A_2 [%]	0	0	8.259228	0
λ_2 [h^{-1}]	0	0	0	0.001769

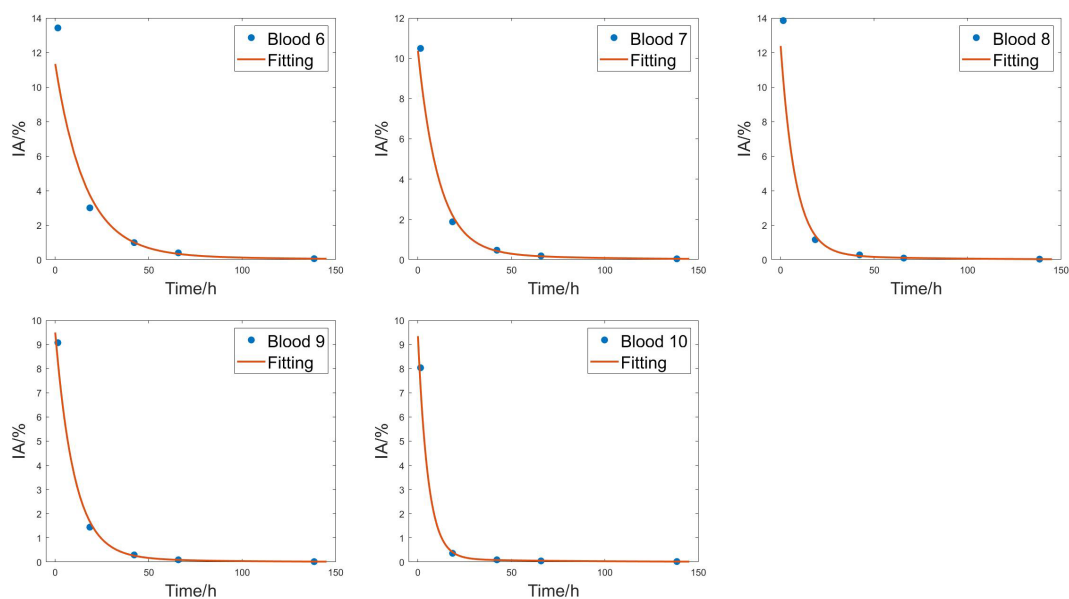


Figure 19. Fitted blood data from five separate patients.

Table 19. Calculated TIACs for blood for patients 6-10 with standard errors (SE).

	Patient 6	Patient 7	Patient 8	Patient 9	Patient 10
TIAC	2.0436	1.3718	1.1528	1.0741	0.5949
SE	1.2787	0.6250	0.4033	0.4992	0.2123

4.4.2. Kidneys

The following exponential functions in Phase 2 were used for fitting of TAC in kidneys data:

$$f_3(t) = [A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t}] e^{-\lambda_{phys} t} \quad (51)$$

$$f_{4d}(t) = A_1 \frac{e^{-\lambda_{phys} t} \{ (1-\alpha) e^{-\lambda_1 t} - \alpha e^{-\lambda_2 t} - (1-2\alpha) e^{-\lambda_{bc} t} \}}{\left\{ \left(\frac{1-\alpha}{\lambda_1 + \lambda_{phys}} \right) - \left(\frac{\alpha}{\lambda_2 + \lambda_{phys}} \right) - \left(\frac{1-2\alpha}{\lambda_{bc} + \lambda_{phys}} \right) \right\}} \quad (52)$$

$$f_{11}(t) = A_1 [\beta e^{-\lambda_1 t} + (1 - \beta) e^{-\lambda_2 t}] e^{-\lambda_{phys} t} \quad (53)$$

where parameter β is the fractional contributions of the corresponding exponentials with values constrained between 0 and 1.

In the case of kidneys, the SP-PBMS was used. The fitting was performed population-wise, including functions with parameters shared between patients. The parameters of the functions were fitted (with different combinations of shared and individual parameters) to the biokinetic data of patients of Phase 1 and 2. Information about the functional shape of the time-activity curve of patients of Phase 1 was used for patients of Phase 2.

The functions below were examined using all biokinetic data of the patient population and a shared parameter approach. The following functions are derived from functions f_3 , f_{4d} and f_{11} with different shared parameters.

$$f_{3S1}(t) = [A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t}] e^{-\lambda_{phys} t} \text{ with shared parameter } A_1 \quad (54)$$

$$f_{3S2}(t) = [A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t}] e^{-\lambda_{phys} t} \text{ with shared parameter } \lambda_1 \quad (55)$$

$$f_{3S3}(t) = [A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t}] e^{-\lambda_{phys} t} \text{ with shared parameters } A_1, \lambda_1 \quad (56)$$

$$f_{3S4}(t) = [A_1 e^{-\lambda_1 t} + A_2 e^{-\lambda_2 t}] e^{-\lambda_{phys} t} \text{ with shared parameters } A_2, \lambda_1, \lambda_2 = 0 \quad (57)$$

$$f_{4dS1}(t) = A_1 \frac{e^{-\lambda_{phys} t} \{ (1-\alpha) e^{-\lambda_1 t} - \alpha e^{-\lambda_2 t} - (1-2\alpha) e^{-\lambda_{bc} t} \}}{\left\{ \left(\frac{1-\alpha}{\lambda_1 + \lambda_{phys}} \right) - \left(\frac{\alpha}{\lambda_2 + \lambda_{phys}} \right) - \left(\frac{1-2\alpha}{\lambda_{bc} + \lambda_{phys}} \right) \right\}} \text{ with shared parameter } A_1 \quad (58)$$

$$f_{4dS2}(t) = A_1 \frac{e^{-\lambda_{phys} t} \{ (1-\alpha) e^{-\lambda_1 t} - \alpha e^{-\lambda_2 t} - (1-2\alpha) e^{-\lambda_{bc} t} \}}{\left\{ \left(\frac{1-\alpha}{\lambda_1 + \lambda_{phys}} \right) - \left(\frac{\alpha}{\lambda_2 + \lambda_{phys}} \right) - \left(\frac{1-2\alpha}{\lambda_{bc} + \lambda_{phys}} \right) \right\}} \text{ with shared parameter } \lambda_1 \quad (59)$$

$$f_{4dS3}(t) = A_1 \frac{e^{-\lambda_{phys} t} \{ (1-\alpha) e^{-\lambda_1 t} - \alpha e^{-\lambda_2 t} - (1-2\alpha) e^{-\lambda_{bc} t} \}}{\left\{ \left(\frac{1-\alpha}{\lambda_1 + \lambda_{phys}} \right) - \left(\frac{\alpha}{\lambda_2 + \lambda_{phys}} \right) - \left(\frac{1-2\alpha}{\lambda_{bc} + \lambda_{phys}} \right) \right\}} \text{ with shared parameter } \alpha \quad (60)$$

$$f_{11S1}(t) = A_1 [\beta e^{-\lambda_1 t} + (1 - \beta) e^{-\lambda_2 t}] e^{-\lambda_{phys} t} \text{ with shared parameter } A_1 \quad (61)$$

$$f_{11S2}(t) = A_1 [\beta e^{-\lambda_1 t} + (1 - \beta) e^{-\lambda_2 t}] e^{-\lambda_{phys} t} \text{ with shared parameter } \beta \quad (62)$$

$$f_{11S3}(t) = A_1 [\beta e^{-\lambda_1 t} + (1 - \beta) e^{-\lambda_2 t}] e^{-\lambda_{phys} t} \text{ with shared parameter } \lambda_1 \quad (63)$$

The index S refers to a shared parameter. The functions f_3 , f_{4d} and f_{11} with unshared parameters were also analyzed. All parameters have been constrained to positive values. All functions were fitted to the biokinetic data of kidneys using the IBMS and the PBMS methods. The fittings were performed using MATLAB and OriginPro.

The function f_{3S4} with only one adjustable parameter A_1 was identified as the function most supported by the kidneys data. Parameters A_2 and λ_1 were shared, they were assumed to be the same for all patients, and the parameter λ_2 was set equal to 0.

Table 20 and Figure 20 show the parameterization and the best fits for kidneys data for 5 patients from Phase 1, performed with the function f_{3S4} . The standard errors for each parameter can also be found in the table. The calculated values for R^2 were in the range between 0.96960 and 0.99761. The results for time-integrated activities are shown in Table 21. The standard errors for TIACs were determined according to the uncertainty propagation described in earlier chapter.

Figure 21 shows the comparison of fitted kidneys data for patient 6-10 performed in Phase 1 and 2 and Figure 22 shows the comparison of calculated TIACs for these patients for kidneys. Patients 10 and 6 show the greatest differences in TIAC values, 11.9 % and 9.4 %, respectively, followed by patient 9 (4.9 %), patient 7 (3.4 %) and patient 8 (2.9 %).

Although the results for the kidneys for patients 6-10 were improved in Phase 2, the standard errors are almost an order of magnitude higher than of the TIACs calculated in Phase 1.

Table 20. Fit parameters for the function f_{3S4} for kidneys data with standard errors (SE), A_2 , λ_1 shared, λ_2 set to 0, AICc = -225.35.

	Patient 6	Patient 7	Patient 8	Patient 9	Patient 10
A_1 [%]	1.588	1.6567	1.4098	3.7578	1.7331
SE	0.2147	0.2485	0.2059	0.5003	0.2328
λ_1 [h^{-1}]	0.01602	0.01602	0.01602	0.01602	0.01602
SE	0.00105	0.00105	0.00105	0.00105	0.00105
A_2 [%]	0.05168	0.05168	0.05168	0.05168	0.05168
SE	0.01197	0.01197	0.01197	0.01197	0.01197
λ_2 [h^{-1}]	0	0	0	0	0
R^2	0.99107	0.98759	0.93793	0.99761	0.96960

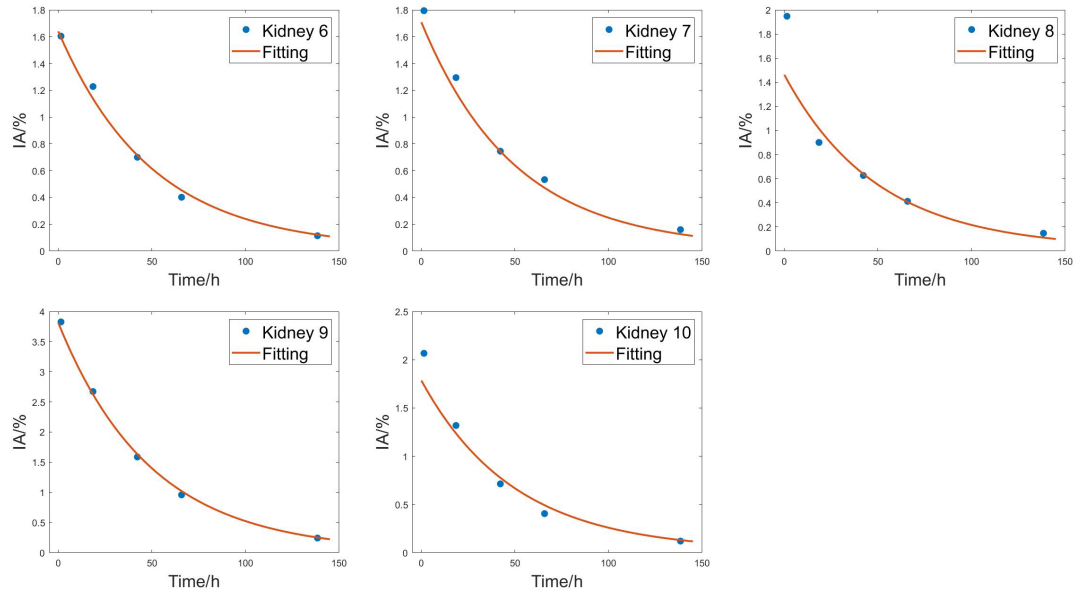


Figure 20. Fitted kidneys data from five separate patients.

Table 21. Calculated TIACs for kidneys for patients 6-10 with standard errors (SE).

	Patient 6	Patient 7	Patient 8	Patient 9	Patient 10
TIAC	0.8987	0.9324	0.8112	1.9642	0.9700
SE	0.1161	0.1319	0.1107	0.2649	0.1255

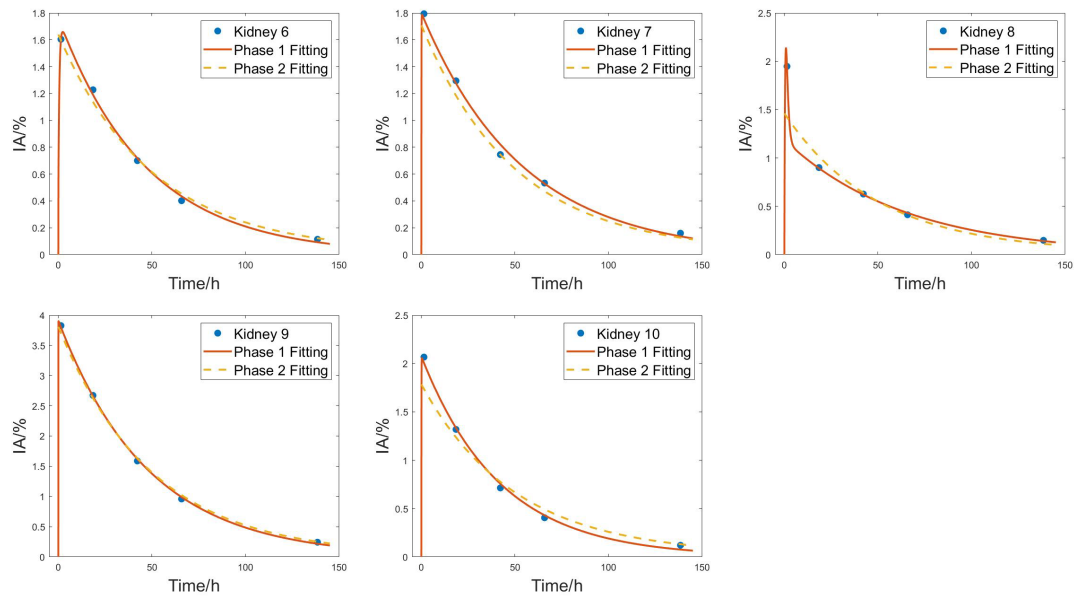


Figure 21. Comparison of the fitted kidneys data in Phase 1 and 2.

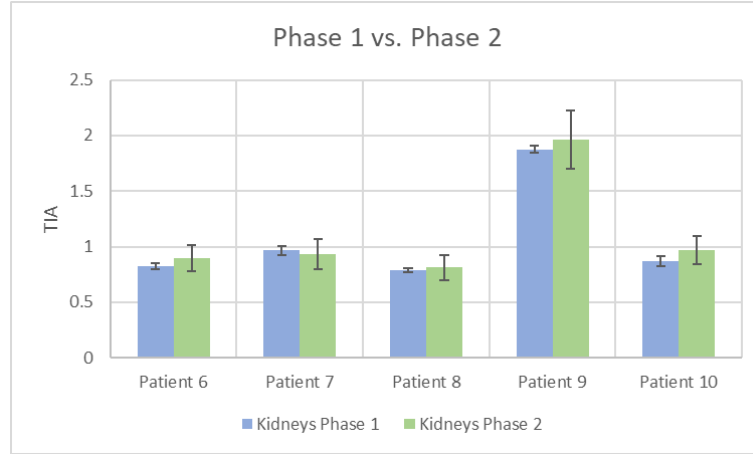


Figure 22. Comparison of calculated TIACs for kidneys for patients 6-10 in Phase 1 and 2.

4.4.3. Tumor

As in Phases 0 and 1, the following exponential function with 4 degree of freedom was used for fitting of TAC in tumor data in Phase 2:

$$f_{4d}(t) = A_1 \frac{e^{-\lambda_{phys}t} \{ (1-\alpha)e^{-\lambda_1 t} - \alpha e^{-\lambda_2 t} - (1-2\alpha)e^{-\lambda_{bc} t} \}}{\left\{ \left(\frac{1-\alpha}{\lambda_1 + \lambda_{phys}} \right) - \left(\frac{\alpha}{\lambda_2 + \lambda_{phys}} \right) - \left(\frac{1-2\alpha}{\lambda_{bc} + \lambda_{phys}} \right) \right\}} \quad (64)$$

As for blood data, the fittings were performed using the non-linear mixed-effects model function *nlmefitsa()* in MATLAB. Also here, the optimization was performed using the *fminsearch* algorithm with a maximum of 100,000 iterations and 100,000 objective function evaluations. The convergence criterion was based on default settings of the algorithm, and the error model used was exponential. The method for approximating the loglikelihood was "importance sampling".

The function was applied to 20 patients of Phase 1 and 2 together to find the best parameter set for function f_{4d} for patients 6-10 of Phase 1. The fixed effect values with standard errors for tumor data are shown in Table 23.

Table 22 and Figure 23 show the parameterization and the best fits for tumor data for 5 patients from Phase 1, performed with the function f_{4d} . The results for time-integrated activities with standard errors are shown in Table 26.

The fixed effect values with standard errors for tumor data are shown in Table 23. The random effects for blood data and the covariance matrix for the random effects are shown in Tables 24-25, respectively. Here, as in the case of blood data, the model defines random effects additively and the individual parameters are computed as a sum of fixed effects (Table 23) and random effects data (Table 24). As in the case of blood data, the standard errors for TIACs were calculated with the variances (diagonal elements) of the covariance matrix of the random effects (Table 25). As shown in Table 25, all the covariances (off-diagonal elements) are equal to 0, which indicates that the parameters are independent.

Table 22. Fit parameters for the function f_{4d} for tumor data, NLME AICc = -706.042.

	Patient 6	Patient 7	Patient 8	Patient 9	Patient 10
A_1 [%]	5.1450	4.8058	10.3068	0.5719	1.4190
α	0.5937	0.5937	0.5937	0.5937	0.5937
λ_1 [h^{-1}]	0.3148	0.0865	1.5658	0.4220	2.3552
λ_2 [h^{-1}]	0.0092	0.0044	0.0038	0.0054	0.0124
λ_{bc} [h^{-1}]	41.589	41.589	41.589	41.589	41.589

Table 23. Fixed effect and standard errors (SE) for tumor data.

	Fixed effect	SE
A_1 [%]	3.2300	0.6436
α	0.5937	0.1063
λ_1 [h^{-1}]	0.3143	0.3934
λ_2 [h^{-1}]	0.0053	0.0010

Table 24. Random effects for tumor data.

	Patient 6	Patient 7	Patient 8	Patient 9	Patient 10
A_1 [%]	1.915023	1.575799	7.076818	-2.65814	-1.81098
α	0	0	0	0	0
λ_1 [h^{-1}]	0.000458	-0.22784	1.251478	0.107714	2.04085
λ_2 [h^{-1}]	0.00397	-0.00083	-0.00144	0.000163	0.007121

Table 25. Covariance matrix for the random effects.

	A_1	α	λ_1	λ_2
A_1 [%]	1.1665	0	0	0
α	0	0	0	0
λ_1 [h^{-1}]	0	0	3.7576	0
λ_2 [h^{-1}]	0	0	0	0.4728

Figure 23. Fitted blood data from five separate patients.

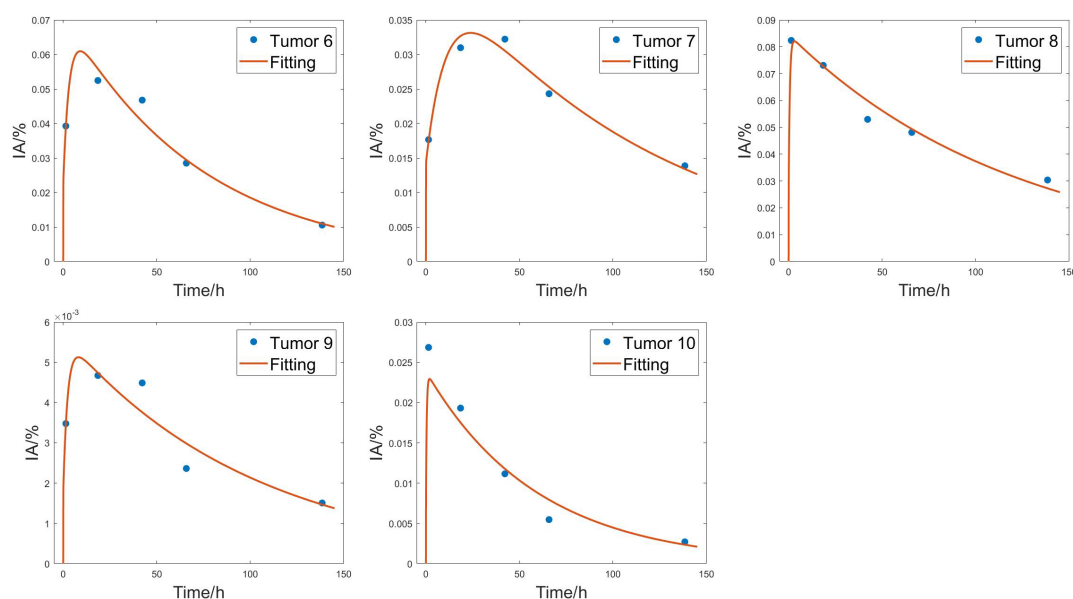


Table 26. Calculated TIACs for tumor for patients 6-10 with standard errors (SE).

	Patient 6	Patient 7	Patient 8	Patient 9	Patient 10
TIAC	0.0515	0.0481	0.1031	0.0057	0.0142
SE	0.0108	0.0108	0.0108	0.0108	0.0108

Figure 24 shows the comparison of fitted tumor data for patient 6-10 performed in Phase 1 and 2 and Figure 25 shows the comparison of calculated TIACs for these patients for tumor. As in the case for kidneys, the greatest difference in TIAC values between Phase 1 and 2 shows the patient 10 (10.1 %). The differences for patients 9, 6, 8 and 7 are 7.5 %, 4.3 %, 3.1 % and 0.2 %, respectively.

In contrast to the kidneys, in the case of tumor, we received no information from the organizers as to whether our results from Phase 2 have improved the results for patients 6-10 from Phase 1 or not. Nevertheless, as for the case with kidneys, the standard errors for TIACs for tumor in Phase 2 are almost an order of magnitude higher than of the TIACs calculated in Phase 1. This indicates a higher level of uncertainty and should be taken into account when analyzing the results.

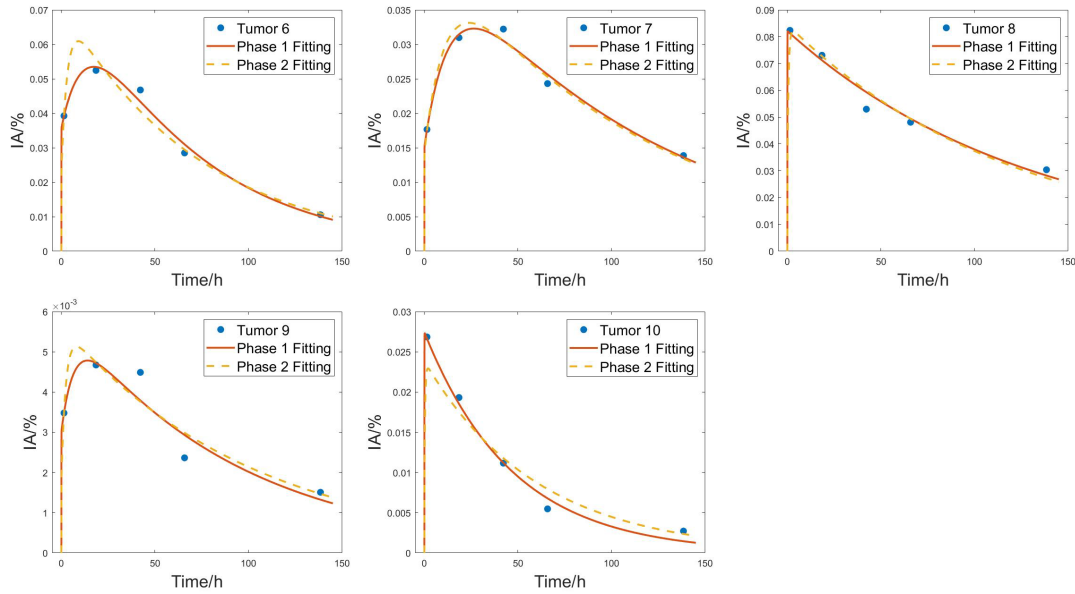


Figure 24. Comparison of the fitted tumor data in Phase 1 and 2.

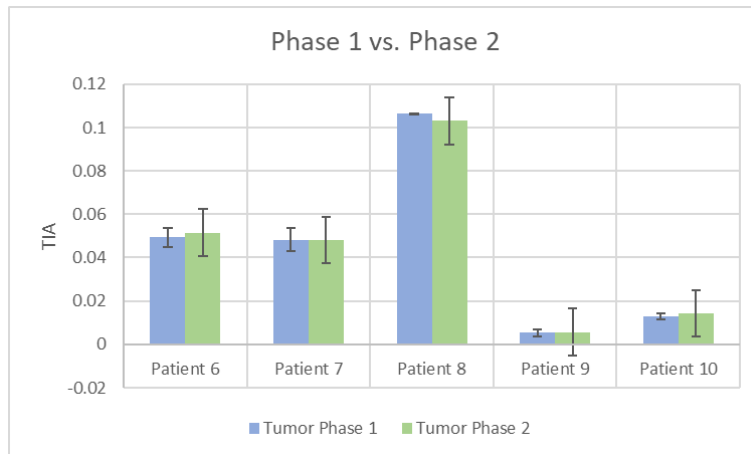


Figure 25. Comparison of calculated TIACs for tumor for patients 6-10 in Phase 1 and 2.

4.5. Discussion

The submissions from Phase 2 were scored and ranked by the Challenge Team according to minimum RMSE and maximum deviation of the results from the ground truth TIAC values of the synthetic data. According to RMSE and maximum deviation of the results from the ground truth TIAC values of the synthetic data, the first place for kidneys with very small RMSE of 0.028 (0.046 in Phase 1) could be achieved. However, in final scoring together with blood and tumor results only 3th place was achieved.

In Phase 2 predefined time-activity data for 15 more patients were provided from organizers to improve the fitting results for patients 6-10 in Phase 1. It was therefore obvious to use population-based methods when fitting the data.

For blood data only one function for fitting purposes was used. The idea of 100 % activity at time point zero and the function f_9 from Phase 1 was rejected and the function f_3 from Phase 0 was used

again. The calculations were performed with MATLAB and OriginPro. In OriginPro the group fitting method was applied to 20 patients of Phase 1 and 2 together to find the best parameter set for function f_3 . Since MATLAB provided better AICs, the results from MATLAB were retained in the end. The reason for the differences in the results between these two tools were certainly different population-based methods but also different fitting algorithms. In case of MATLAB, it was NLME and in case of OriginPro a group fitting in the frame of PBMS. But also, the fitting algorithm itself could cause differences, for example, due to different stopping criteria – what is essential to ensure the convergence when searching the fit function.

In the case of blood data, we cannot speak about improvement of the results in Phase 1, as in Phase 1 a wrong assumption (100 % activity at time point zero) was made, which led to completely incorrect results. Therefore, the TIACs for patients 6-10 from Phase 1 and Phase 2 are incomparable.

Also, for the kidneys data the calculations were performed with MATLAB and OriginPro. Better results could be achieved with shared parameter approach in group fitting with OriginPro than with NLME in MATLAB. The parameters A_2 and λ_1 were shared between 20 patients of Phase 1 and 2 and the start value for parameter A_1 was set as the mean value when applied the IBMS method to all 20 patients. The parameter λ_2 was set to 0. The differences in results between MATLAB and OriginPro were minimal, the AICs were comparable small. Nevertheless, in the case of kidneys data OriginPro delivered slightly better results than MATLAB. Here too, the differences are due to the different population-based methods, as the initial values and others constraints were chosen the same for both MATLAB and OriginPro.

In Phase 0 and 1 for kidneys the function f_{4d} was used when fitting the data. The function has the value 0 at time zero, i.e. the function fulfilled the condition that there is no activity in the kidneys at time zero. However, for function f_3 used in Phase 1 this condition is not fulfilled. For this reason, the calculation of TIACs were performed by integrating of the function from first time point to the infinity and the area between the time zero and the first time point was calculated according to the trapezoidal rule. Thus, in contrast to Phase 1, where the function f_{4d} was found as the best one, in Phase 2 the function f_{3S4} with only one adjustable parameter (two parameters were shared and one fixed) was identified as the function most supported by the kidneys data.

Judging by the good placement for kidneys (first place among all participants with very small RMSE) the function f_{3S4} turned out to be more suitable than function f_{4d} and the parameterization was determined almost perfect. The differences in TIACs between Phase 1 and 2 of more than 11 % resulted in a reduction of RMSE of almost 40 %. It is interesting to note that here the prior knowledge ($f_i(t = 0) = 0$) was not used and data of only 5 given time points were used when fitting the data. The SP-PBMS method with population data without time point 0 provided slightly better results (better AICc) than NLMEMS with population data and 6 time points.

It is also interesting to note the behavior of the fitted curves in Phase 2 when using the population data. As the deviation of the curve from kidneys data for patients 6, 7 and 9 is small, it is quite high for patients 8 and 10. Especially the activity at early time point seems to be much too high. These results might demonstrate that the early time point is less important and can be removed from the fitting process.

For tumor data the best parameter set for the only tested function f_{4d} was found with NLMEMS used in MATLAB. The differences in TIACs between Phase 1 and 2 are also high, more than 10 %, and the highest also for patient 10. However, for tumor data no deviation of the results from the ground truth

TIAC values was provided from organizers. Thus, it is unknown whether the population data from Phase 2 led to improvement or deterioration of TACs and TIACs in Phase 1.

The behavior of the fitted curves in Phase 2 when using the population tumor data is similar to the behavior for kidneys data. Here, too, the greatest deviation can be seen at the early time point. The deviation of the curve for tumor data for patients 7 and 8 is small and it is quite high for patients 6, 9, and 10.

There could be many reasons why the best result in Phase 2 could not be achieved. One of them is the problem with local and global minima. Minimizing the objective function (Eq. 36) depends on minimizing algorithms and can lead to many local minima. Finding an arbitrary local minimum is relatively straightforward by using classical *local optimization* methods but finding the global minimum is far more difficult: analytical methods are frequently not applicable, and the use of numerical solution strategies often leads to very hard challenges. The algorithm for minimizing objective function can deliver different results if the initial values for the fit function parameters were chosen differently and the wrong initial values leads to local minima instead of global minimum. If the distributions of the fit function parameters in the patient population are known, the mean or median values may be used as starting values. But there is no guarantee, that this would lead to the best results. Better approach could be to use search methods for finding optimal starting values for the parameters. This could be the grid or random search method. The grid search is simply an exhaustive searching through a manually specified subset of the hyperparameter space and the random search replaces the exhaustive enumeration of all combinations by selecting them randomly. Unfortunately, such search methods were not used in this challenge due to a lack of time and experience.

The other reason could be the stability check for model selection for different data sets that was not applied in this challenge. The results we got with our sets of parameters may not have been stable and therefore not entirely accurate. There are different methods that could be used to analyze the stability of the model selection, e.g. the Jackknife (Wolter 2007) or Bootstrapping (Efron 1979; Efron and Tibshirani 1993) method. The bootstrapping is the most popular resampling method today. It uses sampling with replacement to estimate the sampling distribution for a desired estimator. The main purpose for this particular method is to evaluate the variance of an estimator. The Jackknife works by sequentially deleting one observation (patient) in the data set, then recomputing the desired statistic. It is computationally simpler than bootstrapping, and more orderly (i.e. the procedural steps are the same over and over again).

Both of them are computationally intensive, so, the lack of time was the main reason why the stability check was not used in this challenge.

Another reason is the number of tested functions. The sets of possible candidates should be large, since the probability of achieving better results increases with the number of functions tested. However, in our case, the sets of possible candidates for blood, kidneys and tumor in all phases were quite small. Only exception were the kidneys data in Phase 2, where 13 different functions or parameter combination was used when fitting the data. Performing the calculations with a large set of function is very time consuming and if there is no suitable software, such as NUKFIT (Kletting et al. 2013), available for this, the calculation is difficult or not possible due to time constraints.

4.6. Conclusion

Calculations in the last phase of the TACTIC grand challenge, in Phase 2, were carried out successfully. The received patient data were successfully used to improve results from Phase 1. Different population-based model selection methods in different tools were used. The results were compared to find the best parameter sets in the fit functions.

The improved time-activity curves for patients from Phase 1 with the help of population-based methods emphasize the importance of such approaches. If the population data is available the population-based model selection methods should be used to obtain better results.

Our calculations with group fitting, we used in OriginPro, or Non-Linear Mixed-Effects model selection from MATLAB, showed similar good results. In our cases, for patient data we received, both population-based methods were similar and could be used, to find the best fit functions and parameter sets.

With the fitted kidneys data, the best result, the 1st place, could be reached. Together with blood and tumor results finally the 3rd place among all participants could be achieved. Many of the theoretical insights gained during the TACTIC grand challenge, like increasing of the number of the functions or stability check, could not be applied due to a lack of time. This is certainly the reason why a very good but not the best result could be achieved. Nevertheless, the knowledge and experience gained in the TACTIC grand challenge was extremely helpful and can be used in the future to get an even better grip on the problem of the parameter finding when fitting the patient data.

5. Lessons learned from the TACTIC Grand Challenge

5.1. Introduction

Throughout the participation in the AAPM TACTIC Grand Challenge, the EURADOS team engaged very deeply, such as at least one virtual meeting every week, with the challenges of time-activity curve (TAC) fitting and time-integrated activity (TIA) estimation under varying data constraints and modeling assumptions. By working through the three phases of the challenge, from the warm-up exercise in Phase 0 through the individualized data TAC-fitting in Phase 1, to population-based modeling in Phase 2, we were able to systematically assess the strengths and limitations of different TAC-fitting and dosimetry approaches. Each phase provided unique insights to participants. In Phase 0, we focused on baseline performance under ideal conditions and learned how to select time point measured data. In Phase 1, we encountered the critical impact of compartment-specific modeling decisions, particularly for blood TACs, this gave us the insight, how the radiopharmaceuticals were injected to the patients and how to model the pharmaceuticals in distribution volumes in the early time. Phase 2 challenged us to incorporate inter-patient variability through population-informed models. In conclusion, these experiences reinforced us to identify essential methodological principles that are not only relevant for synthetic benchmarking, but also directly translatable to clinical and multi-center research contexts. In this section, we summarize the key lessons we learned and outline concrete recommendations for improving the accuracy, robustness, and reproducibility of TAC-based dosimetry workflows. These insights will support ongoing work toward harmonized best practices in radiopharmaceutical therapy dosimetry.

5.2. Lessons we learned and recommendations

5.2.1. Choose the right temporal frequency

In Phase 0 the participation teams had to select between 1 and 5 time points from a synthetic predefined dataset from 1 to 140 hours post injection of [^{177}Lu]Lu-PSMA. Thus, the selection of the appropriate time frequency was one of the exercises in this challenge. In the radiopharmaceutical therapy the radiation dosimetry calculated from multiple-time-point (MTP) imaging, collected with a gamma camera or PET scanner after the radiopharmaceutical is administered to the patient, the selection of appropriate temporal frequency plays a decisive role for an accurate representation of the TAC. The optimal time points at which to measure the activity of an injected radiopharmaceutical for calculating the TACs is a highly topical area of research. The selection of small number of time points or bad distributed time points in MTP would lead to big errors and thus to wrong TACs in the fitting process. Of course, in the medical practice, where the resource intensiveness of MTP dosimetry and the potential inconvenience and discomfort for patients are high, efforts are made to simplify dosimetry image acquisition and the dosimetry calculated from single-timepoint (STP) imaging is aimed at. But when undergoing MTP, the distribution of selected time points should appropriately be chosen and spread over the time interval and cover the uptake and retention pattern of the radiopharmaceutical in each organ. It should be taken into account that for many cases a bi-exponential function might be expected, therefore at least as many data points should be obtained as the number of initially unknown variables in the mathematical curve-fitting function. The MIRD guidelines suggest using at least three time points per exponential term of the fit function.

The effective half-time of the radiopharmaceutical compounds plays a decisive role in selecting an appropriate temporal frequency. As a general rule of thumb, the time points in an interval of about

five times of the half-life should be taken. If a widely used radiopharmaceutical compound is used, like [^{177}Lu]Lu-PSMA the compound for targeted therapy in prostate cancer, the optimum timing regime could be found in the literature. The number and distribution of the time points would result from the uptake and retention pattern of the compound in the organs and tumor. In Phase 0 the time points at 1, 4, 20, 96, and 140 h were finally considered as useful and it was noted that the last time point was important in terms of the accuracy of the tumor TAC determination.

5.2.2. Consider a priori knowledge

5.2.2.1. Consider values at time point zero

The prior knowledge can be very useful and should be integrated into the process when fitting the data. One of the examples of prior knowledge is the consideration of the amount of activity after injected radiopharmaceutical at the time point zero. In our case, for example, from the physiology of the considered organ it can be assumed, that the characteristics of the functions used in the study could be based on the prior knowledge of the biokinetic data in kidneys and tumor were the value of $f_i(t = 0)$ is equal to 0. In this way, an additional point can be added to existing time points, which could make the fitting process more precisely. More possible candidates, i.e. functions with a higher number of parameters become possible. It could then happen that function with 5 parameters and considering the time point 0 provides better results than the function with 4 parameters without considering the time point 0.

Attention should also be paid to the activity at the time point 0 for blood data. Two points concerning the activity at the time point 0 should be considered. Firstly, the volume of distribution. If the volume of distribution is known, it could help to estimate the expected value at time point 0 in blood data. Secondly, the type of injection of activity. Whether the administration is instantaneous or the infusion takes a while. In the case of infusion, the expected value may not be achieved and the TAC could start with 0 at time point 0 (Figure 10).

5.2.2.2. Use an appropriate set of fit functions. Include known function parameters

Another type of prior knowledge are the parameters of the fitting functions. The half-life of the radionuclide used is always known and must be included in the fitting function employed. The half-life should be explicitly included into the equations to avoid possible artefacts when using weighting factors. The best solution can be found easier and faster using a priori knowledge about the parameter values. Adding more information can increase the probability of finding a good fit.

Different constraints could be implemented. For example, the parameter positivity requirement or the physiological conditions like the rate of blood circulation in humans used in function f_{4d} can increase the chances of finding a good fit.

Regarding the population data the knowledge of distribution of the fit function parameters in the patient population could be useful. The population's mean and standard deviations could be considered as prior knowledge and can be integrated into the fitting procedure. In Phase 2 we used fixed or shared parameters, parameters that are similar for all patients, when fitting the data. Such approaches can also be used to improve the results.

In the course of the challenge, we could not identify an approach how to create the best set of fit function and be sure that the best candidate is included. The set of possible candidates for fit functions is to be chosen according to prior knowledge on function parameters. Mono-, bi-, and tri-exponential functions with all possible combination of parameters and factors should be created.

The larger the set, the greater the probability that “right” candidate is included. In the large set a relatively good candidate would be surely found but there is no guarantee that there does not exist a function, that would provide better results. How to find the best set of possible fit function is surely an area of future research.

5.2.2.3. Implement error models

The sources of the uncertainties in the time-activity data are primarily the systematic uncertainty corresponding to the volume determination and quantifications, image noise, registration, patient motion, image reconstruction, scatter and attenuation corrections. If the absolute errors of the time-activity data are known, these errors should be considered when fitting the data. The time-activity data with larger uncertainty could be less weighted than the more accurate and precise data. Also, the influence of the identified outliers could be reduced by appropriate weighting. In our case no information on uncertainty was provided for the time-activity data, neither in Phase 0 nor in Phases 1 and 2. In Phase 0 also no population data was available. In such cases the fit functions could be fitted to the data by a minimum sum of squares algorithm with $1/y$ or $1/y^2$ weighting method. The time-activity data are activity data at different time points, calculated from counts within the volume of interest (VOI) outlined on anatomical or functional imaging data. It can be assumed that Poisson noise of the radionuclide counting is always present. For this reason, the time-activity data without known uncertainty could be weighted at least with $1/y$. The $1/y^2$ weighting to the model could be used if it is assumed, that the errors, for example due to pipetting, are larger than the Poisson noise of the radionuclide counting.

One should also be aware of possible outliers in the data. If, for example, it is known that the probability of errors in the measurements is high, mathematical methods such as Z-score or Interquartile Range (IQR) Method for outlier detection should be used, and the outliers should be excluded or appropriately weighted in the fitting process.

In Phase 2, where the population data were available, the data were used for assigning adequate weights to the data during the fitting process. Also, the error model functions in the frame of population-based-model-selection were used in Phase 2 to estimate the uncertainties from the data. Different type of error models (e.g., additive, proportional or combination of additive and proportional) were implemented and tested.

5.2.3. *Use appropriate integration methods*

There are different methods how to calculate area under the TAC determined from the given time activity data. As shown in the report because of only a few measured time points (possibly not optimally scheduled), numerical integration (trapezoidal rule) may lead to errors. The assumptions made (prior knowledge) regarding uptake or excretion before the first given time point can have a significant impact on the TIAC calculations and should thus be addressed carefully. A further assumption for the time after the last measurement can also lead to errors. Since we used analytical functions in all phases of the challenge to fit the data, it was obvious to determine the area under the curve by integration of the function rather than by trapezoid method. If the first time point is not used in the fitting procedure, the area between the time zero and the first time point can be calculated according to the trapezoidal rule and from first time point to infinity the analytical integration of the fitted function was carried out. In our case for [^{177}Lu]Lu-PSMA the conservative assumption of only physical decay after last time point was not taken into consideration as it was not justified. But, in other cases, for other ligands, it should be considered whether an appearance of an

additional component with long half-life after the last time point makes sense when integrate the time-activity curves.

5.2.4. Use appropriate model selection approach

In different phases of the challenge different model selection methods were used. In Phase 0 no population data were available, for this reason IBMS was applied. In almost all cases functions with at least 4 degrees of freedom were needed to fit 5 given data points. The big disadvantage of IBMS is, when fitting the 5 time points with functions with 4 degrees of freedom the calculation of AICc is not possible. No population information is also taken into account when using IBMS and the uncertainty in calculated TIACs are quite high.

In later phases as population data were available, it could be shown that the use of population-based model selection methods compared with IBMS delivers better results. In Phase 1 PBMS were tested, but not used for final results. In Phase 2 different PBMS were used and compared. It has not been proven that one PBMS method is better than the other. The SP-PBMS, especially for kidneys, provided similar good results as NLMEMS.

The big advantage of the PBMS is the considering of population data. Increased ratio between the number of observed data and the number of estimated parameters reduces uncertainty. Functions with a higher number of parameters become possible.

It is highly recommended to use PBMS, be it SP-PBMS, NLMEMS or NLME-PBMS, if the population data is available.

5.2.5. Consider goodness of fit

Many functions or the same function with different parameterization can fit a given dataset. The following workflow for goodness of fit was determined in the challenge to be useful and is recommended when fitting the data.

The first step is to use only functions with **plausible parameterization** when fitting the data. As shown in Phase 0 and 1, good results can be achieved with parameters which are meaningless from a mathematical, physical or physiological point of view. If the parameters stand for mathematical or physiological quantities, like the fractional contribution of the corresponding exponentials α or the rate of blood circulation in humans λ_{bc} , they should be within the mathematically or physiologically possible range. If the parameters stand for the transfer rates or biological decay, they should all be positive. Another point is the number of used parameters. The model should not contain more parameters than can be justified by the data. The degree of freedom in used fit function should be justified to **avoid overfitting**. With overfitting you would unknowingly extract some of the residual variation (i.e., the noise) as if that variation represented underlying model structure. You should pay close attention to the plausibility of used parameters and avoid any kind of inconsistency at all costs.

The second step is the non-quantitative **visual inspection** of the time-activity curve. The fitted curve should pass through or near most data points and have random trends. This visual test is straightforward to understand and often detects patterns or anomalies in the data. As this test is highly subjective, care should be taken to ensure that no fit functions are incorrectly excluded.

The third step is the calculation of the **coefficient of determination R^2** . This value is between 0 and 1. The good fit corresponds to R^2 close to 1, but a larger R^2 does not necessarily mean a better fit, as the determination of R^2 alone is not sufficient. As in the case of visual inspection, also here care should be taken to ensure that no fit functions are incorrectly excluded because of "small" R^2 .

The first step is to check the **coefficients of variation (CV)** of the fitted parameters of a function. It says how accurately the parameters of the function were determined. It should be not too high. In software packages for data analysis and scientific graphing like OriginPro or SigmaPlot the CV is always given when looking on parameterization of the fit functions. Also, in MATLAB the CV can be checked. The CV of the fitted parameters less than 25 % is considered as precise and if the CV is 50 % it is considered as acceptable. It is recommended to keep the CV of the fitted parameters of a function lower than 50 %.

The fifth step is to check the **correlation matrix (CM)**, the degree of correlation between parameters of the function. The ideal value of the off-diagonal element of CM should be close to zero, what means that the parameters are independent of each other. High correlation tends to widen the confidence intervals. It is recommended that $|CM|$ is smaller than 0.8.

The sixth step is to check whether the **weighted residuals** are randomly distributed or if they show a systematic trend around the zero line. For the latter the model has not captured all information in the data.

The seventh step is to use an **objective criterion to perform the model selection** from many possible candidates. There are different criteria, such as the F-test, BIC or AIC. They represent different theoretical approaches and philosophies. All of these criteria could be used, but especially in the field of molecular radiotherapy, where reality is very complex, often more than two models are compared, and the sample size is small, it is recommended to use **AICc**, the corrected AIC criterion, to perform the model selection. The calculation of AICc should be followed by the determination of **Akaike weights w_i** . The Akaike weights show the probability that the model is the best among the whole set of considered models. The final TIAC should be calculated as weighted sum (w_i as weighting factors) of TIACs with different probabilities.

To analyze the stability of the model selection for different data sets, **stability check methods** such as **Jackknife or Bootstrapping** can be used. Unfortunately, the stability check was not applied in this challenge. The stability check methods are computationally intensive, so, the lack of time was the main reason why these methods were not used. Nevertheless, in the fitting procedures the stability check methods could be decisive to achieve more accurate results and should be applied if possible.

The eighth step is to specify the **uncertainty of the calculated TIACs**. The calculated uncertainty of the TIAC value should be justified and verified. In our case we used analytical uncertainty propagation method to determine standard errors in the TIAC values. However, if more information is available, e.g. on the type of probability distribution based on known uncertainties for each parameter, the use of Monte Carlo simulation for uncertainty analysis would be more useful. The output distribution, from which statistical measures such as mean, standard deviation, and confidence intervals can be derived, would provide a deeper insight into the uncertainty of the calculated values.

5.3. Conclusion

Our EURADOS team demonstrated strong performance and gained valuable insight across all three phases of the AAPM TACTIC Grand Challenge, underscoring both our technical capabilities and our commitment to advancing dosimetry in radiopharmaceutical therapy. In the initial Phase 0, we gained experiences of individualized TAC fitting workflow under unconstrained conditions, this affirmed our readiness to tackle more demanding challenges in TAC curve fitting for radiopharmaceuticals in future projects. Transitioning to Phase 1, where each patient's data was

modeled in isolation, we encountered the critical importance of compartment-specific strategies, particularly for blood TACs. This phase pinpointed areas for refinement in early time weighting, model selection, and uncertainty management, leading to targeted improvements in our workflow. Building on these lessons, in Phase 2 we can integrate population-based modeling techniques to balance individual data reliability with population kinetic patterns. By employing non-linear mixed-effects and Bayesian hierarchical approaches on a synthetic cohort of 15 patients, we strengthened our ability to generalize across heterogeneous data and leverage inter-patient information. This achievement emphasized the tangible benefits of population priors in mitigating sparse sampling and enhancing TIAC precision, while also revealing the trade-offs between model complexity, interpretability, and computational tractability.

Through this challenge, EURADOS gained deep appreciation for:

- Model robustness - The necessity of flexible, compartment-like models (e.g., multi-term exponentials for blood) and adaptive weighting schemes to gain the pharmacokinetics.
- Uncertainty quantification - The impact of uncertainty and limited time points on TIAC accuracy, and the value of incorporating formal uncertainty quantification to guide model confidence and subsequent dosimetry assessment.
- Population-based workflows - The power of hierarchical and data-driven methods to stabilize fits under real-world constraints, highlighting how shared priors can augment individualized estimates without sacrificing personalization.
- Best practices - The importance of standardized evaluation frameworks for transparent comparison and method harmonization across research groups.

In a prospective conclusion, the insights gleaned from our participation will lead EURADOS ongoing efforts to refine pharmacokinetic modeling approaches and to establish consensus guidelines for TAC fitting in clinical RPT dosimetry. We plan to use our experiences learned in this challenge and collaborate with the broader community to integrate the modern machine learning-driven uncertainty estimation. Finally, the TACTIC Grand Challenge has not only validated our technical approach but has also introduced new directions for research, education, and clinical translation in personalized dosimetry.

6. References

- Abuqbeith M, Demir M, Uslu-Bešli L, et al (2018) Blood clearance and occupational exposure for ¹⁷⁷Lu-DOTATATE compared to ¹⁷⁷Lu-PSMA radionuclide therapy. *Radiat Environ Biophys* 57:55–61. <https://doi.org/10.1007/s00411-017-0721-6>
- Akaike H (1974) A new look at the statistical model identification. *IEEE Trans Automat Contr* 19:716–723. <https://doi.org/10.1109/TAC.1974.1100705>
- Akaike H (1998) Information Theory and an Extension of the Maximum Likelihood Principle. In: Parzen E, Tanabe K, Kitagawa G (eds) *Selected Papers of Hirotugu Akaike*. Springer New York, New York, NY, pp 199–213
- Barrett PHR, Bell BM, Cobelli C, et al (1998) SAAM II: Simulation, analysis, and modeling software for tracer and pharmacokinetic studies. *Metabolism* 47:484–492. [https://doi.org/10.1016/S0026-0495\(98\)90064-6](https://doi.org/10.1016/S0026-0495(98)90064-6)
- Burnham KP, Anderson DR (eds) (2002) *Model Selection and Multimodel Inference*, 2nd edn. Springer New York, New York, NY
- Burnham KP, Anderson DR, Burnham KP (2002) *Model selection and multimodel inference: a practical information-theoretic approach*, 2nd ed. Springer, New York
- Carpanese D, Ferro-Flores G, Ocampo-Garcia B, et al (2020) Development of ¹⁷⁷Lu-scFvD2B as a Potential Immunotheranostic Agent for Tumors Overexpressing the Prostate Specific Membrane Antigen. *Sci Rep* 10:9313. <https://doi.org/10.1038/s41598-020-66285-2>
- Cytawa W, Hendel R, Tomasik B, et al (2023) Early biochemical and radiographic response after one cycle of [¹⁷⁷Lu]Lu-PSMA I&T radioligand therapy in metastatic castration-resistant prostate cancer patients. *Eur J Nucl Med Mol Imaging* 50:3765–3776. <https://doi.org/10.1007/s00259-023-06326-w>
- Efron B (1979) Bootstrap Methods: Another Look at the Jackknife. *Ann Statist* 7:. <https://doi.org/10.1214/aos/1176344552>
- Efron B, Tibshirani RJ (1993) *An Introduction to the Bootstrap*. Springer US, Boston, MA
- Gear JJ, Cox MG, Gustafsson J, et al (2018) EANM practical guidance on uncertainty analysis for molecular radiotherapy absorbed dose calculations. *Eur J Nucl Med Mol Imaging* 45:2456–2474. <https://doi.org/10.1007/s00259-018-4136-7>
- Glatting G, Kletting P, Reske SN, et al (2007) Choosing the optimal fit function: Comparison of the Akaike information criterion and the F-test. *Medical Physics* 34:4285–4292. <https://doi.org/10.1118/1.2794176>
- Glatting G, Reske SN (1999) Treatment of radioactive decay in pharmacokinetic modeling: Influence on parameter estimation in cardiac ¹³N-PET. *Med Phys* 26:616–621. <https://doi.org/10.1118/1.598561>
- Hardiansyah D, Riana A, Beer AJ, Glatting G (2023) Single-time-point dosimetry using model selection and nonlinear mixed-effects modelling: a proof of concept. *EJNMMI Phys* 10:12. <https://doi.org/10.1186/s40658-023-00530-1>

- Hardiansyah D, Riana A, Kletting P, et al (2021) A population-based method to determine the time-integrated activity in molecular radiotherapy. *EJNMMI Phys* 8:82. <https://doi.org/10.1186/s40658-021-00427-x>
- Hennrich U, Eder M (2022) [177Lu]Lu-PSMA-617 (PluvictoTM): The First FDA-Approved Radiotherapeutical for Treatment of Prostate Cancer. *Pharmaceuticals* 15:1292. <https://doi.org/10.3390/ph15101292>
- International Commission on Radiological Protection (ICRP) (2019) Occupational intakes of radionuclides: part 4. ICRP Publication 141. *Ann ICRP* 48:1–514
- Ivashchenko OV, O'Doherty J, Hardiansyah D, et al (2024) Time-Activity data fitting in molecular Radiotherapy: Methodology and pitfalls. *Physica Medica* 117:103192. <https://doi.org/10.1016/j.ejmp.2023.103192>
- Jackson PA, Hofman MS, Hicks RJ, et al (2020) Radiation dosimetry in ¹⁷⁷Lu-PSMA-617 therapy using a single posttreatment SPECT/CT scan: A novel methodology to generate time- and tissue-specific dose factors. *J Nucl Med* 61:1030–1036. <https://doi.org/10.2967/jnumed.119.233411>
- Jacquez JA (1996) Compartmental analysis in biology and medicine, 3rd Edition. BioMedware, Ann Arbor, MI
- Kabasakal L, Toklu T, Yeyin N, et al (2017) Lu-177-PSMA-617 Prostate-Specific Membrane Antigen Inhibitor Therapy in Patients with Castration-Resistant Prostate Cancer: Stability, Bio-distribution and Dosimetry. *Mirt* 26:62–68. <https://doi.org/10.4274/mirt.08760>
- Khawar A, Eppard E, Roesch F, et al (2019) Biodistribution and post-therapy dosimetric analysis of [177Lu]Lu-DOTAZOL in patients with osteoblastic metastases: first results. *EJNMMI Res* 9:102. <https://doi.org/10.1186/s13550-019-0566-x>
- Kletting P, Kull T, Reske SN, Glatting G (2009) Comparing time activity curves using the Akaike information criterion. *Phys Med Biol* 54:N501–N507. <https://doi.org/10.1088/0031-9155/54/21/N01>
- Kletting P, Schimmel S, Kestler HA, et al (2013) Molecular radiotherapy: The NUKFIT software for calculating the time-integrated activity coefficient. *Medical Physics* 40:102504. <https://doi.org/10.1118/1.4820367>
- Kletting P, Schuchardt C, Kulkarni HR, et al (2016) Investigating the Effect of Ligand Amount and Injected Therapeutic Activity: A Simulation Study for ¹⁷⁷Lu-Labeled PSMA-Targeting Peptides. *PLoS ONE* 11:e0162303. <https://doi.org/10.1371/journal.pone.0162303>
- Konijnenberg M (2011) From imaging to dosimetry and biological effects. *Q J Nucl Med Mol Imaging* 55:44–56
- Kratochwil C, Giesel FL, Stefanova M, et al (2016) PSMA-targeted radionuclide therapy of metastatic castration-resistant prostate cancer with ¹⁷⁷Lu-labeled PSMA-617. *J Nucl Med* 57:1170–1176. <https://doi.org/10.2967/jnumed.115.171397>
- Lassmann M, Hänscheid H, Chiesa C, et al (2008) EANM Dosimetry Committee series on standard operational procedures for pre-therapeutic dosimetry I: blood and bone marrow dosimetry in differentiated thyroid cancer therapy. *Eur J Nucl Med Mol Imaging* 35:1405–1412. <https://doi.org/10.1007/s00259-008-0761-x>
- Lomax RG (2007) Statistical concepts: a second course, 3. ed. Lawrence Erlbaum Assoc, Mahwah, NJ

- McClave JT, Sincich TT (2016) Statistics. Pearson Education
- Muzic RF, Christian BT (2006) Evaluation of objective functions for estimation of kinetic parameters. *Medical Physics* 33:342–353. <https://doi.org/10.1118/1.2135907>
- Novartis Pharmaceuticals Corporation (2025) PLUVICTO: Highlights of prescribing information. https://www.accessdata.fda.gov/drugsatfda_docs/label/2025/215833s021s024lbl.pdf. Accessed 22 May 2025
- Perazzolo S (2024) SAAM II : A general mathematical modeling rapid prototyping environment. *CPT Pharmacom & Syst Pharma* 13:1088–1102. <https://doi.org/10.1002/psp4.13181>
- Peters SMB, Mink MCT, Privé BM, et al (2023) Optimization of the radiation dosimetry protocol in Lutetium-177-PSMA therapy: toward clinical implementation. *EJNMMI Res* 13:6. <https://doi.org/10.1186/s13550-023-00952-z>
- Rinscheid A, Kletting P, Eiber M, et al (2020) Influence of sampling schedules on [177Lu]Lu-PSMA dosimetry. *EJNMMI Phys* 7:41. <https://doi.org/10.1186/s40658-020-00311-0>
- Sachs L (1999) *Angewandte Statistik*. Springer Berlin Heidelberg, Berlin, Heidelberg
- Santos-Cuevas C, Ferro-Flores G, García-Pérez FO, et al (2018) ¹⁷⁷ Lu-DOTA-HYNIC-Lys(Nal)-Urea-Glu: Biokinetics, Dosimetry, and Evaluation in Patients with Advanced Prostate Cancer. *Contrast Media & Molecular Imaging* 2018:1–10. <https://doi.org/10.1155/2018/5247153>
- Schwarz G (1978) Estimating the dimension of a model. *Ann Statist* 6:. <https://doi.org/10.1214/aos/1176344136>
- Siegel JA, Thomas SR, Stubbs JB, et al (1999) MIRD Pamphlet No. 16: Techniques for quantitative radiopharmaceutical biodistribution data acquisition and analysis for use in human radiation dose estimates. *J Nucl Med* 40:375–615
- Sjögreen Gleisner K, Chouin N, Gabina PM, et al (2022) EANM dosimetry committee recommendations for dosimetry of ¹⁷⁷Lu-labelled somatostatin-receptor- and PSMA-targeting ligands. *Eur J Nucl Med Mol Imaging* 49:1778–1809. <https://doi.org/10.1007/s00259-022-05727-7>
- Stoica P, Selen Y (2004) Model-order selection. *IEEE Signal Process Mag* 21:36–47. <https://doi.org/10.1109/MSP.2004.1311138>
- Strand S, Zanzonico P, Johnson TK (1993) Pharmacokinetic modeling. *Medical Physics* 20:515–527. <https://doi.org/10.1118/1.597047>
- Wang Z, Mårtensson L, Nilsson R, et al (2005) Blood pharmacokinetics of various monoclonal antibodies labeled with a new trifunctional chelating reagent for simultaneous conjugation with 1,4,7,10-tetraazacyclododecane- *N,N'*, *N''*, *N'''*-tetraacetic acid and biotin before radiolabeling. *Clinical Cancer Research* 11:7171s–7177s. <https://doi.org/10.1158/1078-0432.CCR-1004-0001>
- Williams LE, Odom-Maryon TL, Liu A, et al (1995) On the correction for radioactive decay in pharmacokinetic modeling. *Medical Physics* 22:1619–1626. <https://doi.org/10.1118/1.597421>
- Wolter KM (2007) The Jackknife Method. In: *Introduction to Variance Estimation*. Springer New York, New York, NY, pp 151–193

- Zang J, Fan X, Wang H, et al (2019) First-in-human study of ^{177}Lu -EB-PSMA-617 in patients with metastatic castration-resistant prostate cancer. *Eur J Nucl Med Mol Imaging* 46:148–158. <https://doi.org/10.1007/s00259-018-4096-y>
- Zhang Z, Cheng Y, Liu NC (2014) Comparison of the effect of mean-based method and z-score for field normalization of citations at the level of Web of Science subject categories. *Scientometrics* 101:1679–1693. <https://doi.org/10.1007/s11192-014-1294-7>