



Analysis methodology and development of a statistical tool for biodistribution data from internal contamination with actinides

Stephanie Lamart, Nina Griffiths, Nicolas Tchitchek, Jaime Angulo, Anne van der Meeren

► To cite this version:

Stephanie Lamart, Nina Griffiths, Nicolas Tchitchek, Jaime Angulo, Anne van der Meeren. Analysis methodology and development of a statistical tool for biodistribution data from internal contamination with actinides. *Journal of Radiological Protection*, 2017, 37 (1), pp.296-308. 10.1088/1361-6498/37/1/296 . hal-03856452

HAL Id: hal-03856452

<https://hal.science/hal-03856452>

Submitted on 16 Nov 2022

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

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



Distributed under a Creative Commons Attribution 4.0 International License

Practical Matter

Analysis methodology and development of a statistical tool for biodistribution data from internal contamination with actinides

Stephanie Lamart¹, Nina M Griffiths¹, Nicolas Tchitchek²,
Jaime F Angulo¹ and Anne Van der Meeren¹

¹ Laboratoire de RadioToxicologie, CEA, Université Paris-Saclay,
91297 Arpajon, France

² Immunology of viral infections and autoimmune diseases, CEA—Université
Paris Sud 11—INSERM U1184, Fontenay-aux-Roses, France

E-mail: stephanie.lamart@cea.fr

Received 12 May 2016, revised 25 October 2016

Accepted for publication 1 November 2016

Published 28 February 2017



Abstract

The aim of this work was to develop a computational tool that integrates several statistical analysis features for biodistribution data from internal contamination experiments. These data represent actinide levels in biological compartments as a function of time and are derived from activity measurements in tissues and excreta. These experiments aim at assessing the influence of different contamination conditions (e.g. intake route or radioelement) on the biological behavior of the contaminant. The ever increasing number of datasets and diversity of experimental conditions make the handling and analysis of biodistribution data difficult. This work sought to facilitate the statistical analysis of a large number of datasets and the comparison of results from diverse experimental conditions. Functional modules were developed using the open-source programming language *R* to facilitate specific operations: descriptive statistics, visual comparison, curve fitting, and implementation of biokinetic models. In addition, the structure of the datasets was harmonized using the same table format. Analysis outputs can be written in text files and updated data can be written in the consistent table format. Hence, a data repository is built progressively, which is essential for the optimal use of animal data. Graphical representations can be automatically generated and saved as image files. The resulting computational tool was applied using data derived from wound contamination experiments conducted under different conditions.



Original content from this work may be used under the terms of the [Creative Commons Attribution 3.0 licence](https://creativecommons.org/licenses/by/3.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

In facilitating biodistribution data handling and statistical analyses, this computational tool ensures faster analyses and a better reproducibility compared with the use of multiple office software applications. Furthermore, re-analysis of archival data and comparison of data from different sources is made much easier. Hence this tool will help to understand better the influence of contamination characteristics on actinide biokinetics. Our approach can aid the optimization of treatment protocols and therefore contribute to the improvement of the medical response after internal contamination with actinides.

Keywords: internal contamination, radionuclides, statistical analysis, biokinetics, biodistribution data, computational tool, radiotoxicology

(Some figures may appear in colour only in the online journal)

1. Introduction

Understanding of the behaviour of actinides within the body after internal contamination is essential to the accurate assessment of doses, and to the development of optimal medical treatment of contaminated personnel (ICRP 1993, 2015, Castellani *et al* 2013). Actinides can be transported from the site of entry in the body to systemic tissues, be retained in different biological compartments and excreted more or less quickly. This behavior referred to as biokinetics depends largely on the nature of the contamination (e.g. the route of entry, the radioelement(s) and the physico-chemical form of the compound). Animal experiments have been used to evaluate the influence of several contamination conditions on actinide biokinetics. In these experiments, the radiocontaminant is administered through one possible route of intake (e.g. inhalation or wound). The actinide amount in the different biological compartments is monitored as a function of time using activity measurements in excreta and in tissues collected after sequential euthanasia. These measurements of actinide biodistribution generated a large number of data depending on the conditions of contamination (Métivier *et al* 1989, Durbin 2006, Van der Meeren *et al* 2008, 2012, Griffiths *et al* 2012, Doyle-Eisele *et al* 2014, Weber *et al* 2014).

However, analysis of biodistribution data can be complicated due to the large number of datasets and the diversity of the experimental conditions explored. Furthermore, this analysis requires several statistical techniques and should be well documented. Hence, it has been challenging to find a single tool that meets all these requirements and that would facilitate: (i) the re-analysis of multiple historical datasets to test new hypotheses as an alternative to animal experiments, (ii) the comparison of results across experiments, and (iii) the comparison of results with human data. Furthermore, the lack of such a tool may prevent comprehensive data analysis and the development of new knowledge.

To overcome these difficulties, this work proposed a common organization for biodistribution data and a harmonization of analysis methods through the development of an integrated computational tool. The required functional specifications are the following. The standard data format should enable archiving of data from previous experiments, and also inclusion of new data (Schofield *et al* 2010). The computational tool that takes these formatted data as input should facilitate biodistribution data analyses. These analyses aim mainly at evaluating the influence of diverse experimental conditions on activity retention and excretion as a function of time after contamination (Durbin 2006). To conduct these analyses, descriptive statistics is generally and primarily applied: the central tendency and spread of the data are calculated at each time point for each group of animals exposed to the same experimental

conditions (e.g. group mean and standard deviation). To get a direct understanding of actinide biokinetics, it is often necessary to represent as a function of time the central values of activities retained in tissues or excreted. Furthermore, to quantify the kinetics of activity retention and excretion, the biodistribution data are fitted using a sum of exponential functions of the time (Rang *et al* 1995, Doyle-Eisele *et al* 2014, Weber *et al* 2014). Also, when a comprehensive set of data is available, biokinetic models can be derived to predict actinide biokinetics depending on the characteristics of the contamination (Leggett 1992, Leggett *et al* 2005). In internal dosimetry, biokinetic models are compartmental models classified in two types: (i) entry models describing the radioelement behavior from the site of contamination to the blood, e.g. respiratory or wound models; and (ii) systemic models describing the radioelement uptake and clearance in systemic tissues, and the urinary and fecal excretions. Biokinetic models are subsequently used in the assessment of internal dose after contamination (ICRP 1993, 2015, Castellani *et al* 2013). Therefore, the comparison of biodistribution data with the predictions of biokinetic models and with human data is essential to consolidate and improve further the understanding of influential parameters in the biokinetics of actinides.

To include the necessary features for a comprehensive statistical analysis of biodistribution data, the computational tool presented in this paper was implemented using the open-source *R Project for Statistical Computing* (R Core Team 2015). The *R* language has powerful graphics capabilities and, along with *RStudio* (RStudio Team 2015), presents a convenient environment for development. A practical example of data derived from a set of wound contamination experiments is used as an illustration along this paper (Griffiths *et al* 2012).

2. Analysis methodology and development of the computational tool

2.1. Formatting and archiving the raw data

Biodistribution data comprise radionuclide activities (Bq) measured in tissues and excreta at several time points for different groups of animals exposed to different contamination conditions. The approach presented in this paper was based on an example of biodistribution data derived from an experimental rodent model of wound contamination (Griffiths *et al* 2012). The purpose of this experimental setting was to quantify the influence of the experimental conditions on the biodistribution data by comparing the resulting excretion and retention data between wound types, radioelements and physico-chemical forms. The experimental conditions of interest included two realistic wounds (shallow or deep incision), two radioelements (plutonium or americium), and two physico-chemical forms (nitrate or oxide). The oxide was a mixed oxide, MOX, containing a mixture of uranium, plutonium (Pu) and americium (Am). Only the behaviour of Pu and Am were considered since these elements are the major alpha emitters in MOX that contribute to a potential radiological risk following contamination (Priest 2001). The deposited activity was measured locally at the wound site at day 0 immediately after contamination when the animal was still anesthetized, using a sodium iodide scintillation detector and a proportional counter (Griffiths *et al* 2012). The daily urinary excretion was measured from day 1 to 7 and the skeleton and liver activity retentions were measured at day 7. There were five or six animals per set of experimental conditions.

One difficulty when analyzing biodistribution data is to handle the heterogeneity of data format. To develop a global approach for the statistical analysis of biodistribution data, it was required to define a standardized format of table. This table format should primarily contain the radioelement activity measured in tissue or excreta samples at different time points. It should also include the administered activity and the initial proportion of each radioelement for contaminants composed of a mixture of radioelements like the MOX. The content of this

Table 1. List and description of the variables used in the standard format for the raw data file.

Number	Variable	Description
1	original_data_file	Name of the file where the data were originally stored
2	exp_id	Identification of the experiment
3	rat_id	Identification of the animal
4	solution_id	Identification of the solution
5	pH	pH of the solution
6	chemical_form	Chemical form of the contaminant
7	element	Chemical element measured
8	isotope	Isotope measured
9	element_alpha_act_fraction	Activity fraction of the element in the contaminant
10	intake_route	Route of intake
11	contamination_date	Date of contamination
12	adm_alpha_act_Bq	Total administered activity in becquerels
13	excreta_or_tissue	Nature of the sample
14	time_h	Sample collection date in hours after contamination
15	time_d	Sample collection date in days after contamination
16	act_Bq	Measured activity in the sample in becquerels

table should facilitate (i) the assessment of the influence of any contamination conditions; (ii) the data comparison between experimental conditions and with human data; and (iii) the re-analysis of data to test new hypotheses. It was therefore necessary to include in this table a detailed identification of the sample and a full description of the contamination conditions.

In the format proposed in this methodology, data were structured in tables as comma separated values (CSV) files, in which the header contained variable names (e.g. route of intake, element and physico-chemical form) and each line contained the value of each variable for one observation (e.g. deep wound, americium and nitrate, respectively). The full list of variables and the corresponding description can be found in table 1. Each observation corresponded to one activity measurement of a sample conducted at one time point under a given set of experimental conditions. One CSV file was created per group of animals exposed to the same set of experimental conditions and for one sample type (e.g. urine). Raw data from the wound experiments were stored using this standard format and directly imported into *R* (using *dataframe objects*). There were 42 CSV files, corresponding to the different experimental conditions, which consisted of a total of 900 observations. Table 2 illustrates the standard format on a selected set of data corresponding to wound contamination.

2.2. Implementing the biodistribution data analysis in a single computational tool

The computational tool was designed to handle and analyze each raw data file individually, i.e. for each group of animals exposed to the same experimental conditions and for one sample type (e.g. urine). Programs were written in *R* (*R scripts*) to automate the specific operations related to data analysis. The tool was subdivided into four modules: descriptive statistics, regression analysis, biokinetic modeling and data comparison (figure 1). The computational tool was also designed to build a structured repository of data and analyses.

Table 2. Example of raw biodistribution data from the wound experiments organized in the table format. Each block of this table represents one time point with six animals per group. One line of this table represents one observation (i.e. activity measurement in one biological sample in a given set of experimental conditions). Each column represents one variable characterizing the samples.

rat_id	chemical_form	element	element_alpha_ act_fraction	adm_alpha_ act_Bq	time_d	act_Bq
MB_101_51	MOX	Pu	0.73	10293	1	0.4
MB_101_52	MOX	Pu	0.73	7618	1	0.8
MB_101_53	MOX	Pu	0.73	12209	1	6.3
MB_101_54	MOX	Pu	0.73	7412	1	4.1
MB_101_55	MOX	Pu	0.73	14545	1	5.8
MB_101_56	MOX	Pu	0.73	6417	1	0.7
MB_101_51	MOX	Pu	0.73	10293	2	1.4
MB_101_52	MOX	Pu	0.73	7618	2	0.4
MB_101_53	MOX	Pu	0.73	12209	2	2.3
MB_101_54	MOX	Pu	0.73	7412	2	0.8
MB_101_55	MOX	Pu	0.73	14545	2	4.6
MB_101_56	MOX	Pu	0.73	6417	2	3.3
AB_111_11	Nitrate	Am	1	6348	3	28.4
AB_111_12	Nitrate	Am	1	5827	3	20.6
AB_111_13	Nitrate	Am	1	4464	3	15.7
AB_111_14	Nitrate	Am	1	4753	3	16.8
AB_111_15	Nitrate	Am	1	6943	3	37.5
AB_111_16	Nitrate	Am	1	6274	3	24.4
AB_111_11	Nitrate	Am	1	6348	4	19.0
AB_111_12	Nitrate	Am	1	5827	4	21.0
AB_111_13	Nitrate	Am	1	4464	4	17.7
AB_111_14	Nitrate	Am	1	4753	4	9.6
AB_111_15	Nitrate	Am	1	6943	4	28.2
AB_111_16	Nitrate	Am	1	6274	4	21.0

Hence, each step was implemented so that graphical representations of the data and results were automatically generated. In addition, updated data and results of analysis were saved in text files. Several rules were observed to name the files: no blanks, no special characters (e.g. accents), and inclusion of the experiment identification and of the nature of the data (e.g. raw or statistics). The naming of output files and the storage in the appropriate folder was done automatically. The comparison of results between experimental conditions was also implemented. The following sections will deal in turn with each module of the computational tool.

2.3. Module for descriptive statistics

A functionality was developed to analyze each raw data file, (i.e. set of experimental conditions with one sample type; Module #1 in figure 1). Because the administered activity may vary between experiments of internal contamination (table 2), it is often useful to normalize the raw data to the administered activity. Data normalization enables comparing data between various experiments. Furthermore, a radioelement may behave differently in the body if administered as a single contaminant or if administered together with other elements. In case of a mixture of radioelements (e.g. plutonium and americium), it is therefore necessary to follow-up the

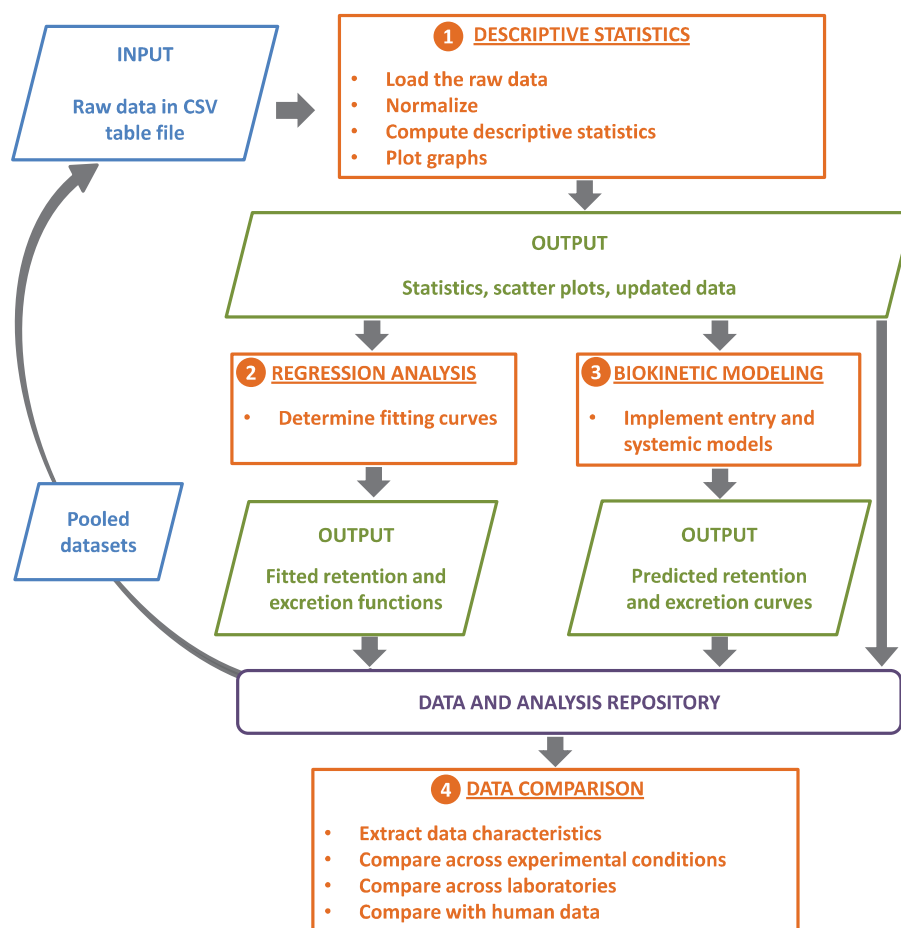


Figure 1. Flow chart of the functionalities of the computational tool developed to conduct statistical analyses of biodistribution data. The four main modules are described in the text.

percentage of the activity due to a particular element in relation to the total activity. Hence the automatic computation of new variables corresponding to different normalizations was included in this module. These variables were: (i) the percentage of the total activity in the sample per unit of administered activity, (ii) the percentage of the element activity in the sample per unit of administered activity, and (iii) the percentage of activity from one element in relation to the total activity in the sample.

The analysis of biodistribution data necessitates the determination of the central tendencies. Group statistics were systematically computed for each set of experimental conditions and time point (Freedman 2007). The following group statistics were derived for items (ii) and (iii) described above: the mean and standard deviation, the median, the extrema, and the standard error of the mean. The number of animals per dataset was also retrieved in the process.

To follow-up retention and excretion of activity as a function of time, functionalities for automatic plotting were developed thus enabling a direct, fast and convenient visual

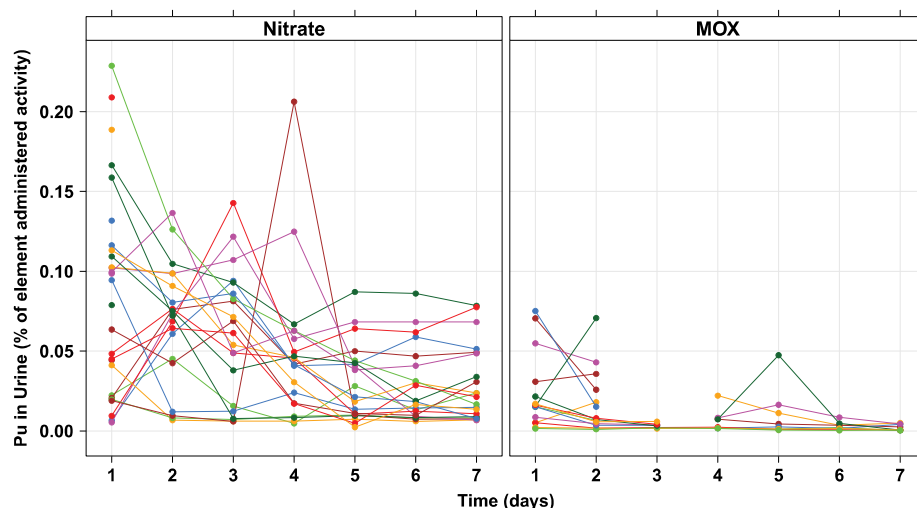


Figure 2. Comparison of urinary excretion of plutonium after a deep wound between two physico-chemical forms of plutonium: nitrate and oxide (MOX: mixed oxide of uranium, plutonium and americium) using scatter plots. Each line represents one animal. Broken lines indicate missing values.

representation of the data. The plotting of raw data and group statistics were implemented using scatter plots (Sarkar 2008, 2015). Biodistribution data of actinides are usually characterized by a wide variability between animals. In addition, the complexity of actinide manipulation, animal experimentation and actinide spectrometry may lead to the generation of extreme values. Identification of extremes is necessary to trace back the information related to the corresponding sample, to understand potential reasons and to decide how extremes should be treated. Therefore, it was essential to represent the (normalized) raw data, before calculating mean values, and to visualize directly the data spread at each time point easily. Finally, a qualitative comparison of data across different experimental conditions was facilitated using side by side representation of data in different panels with the same axis scales. As an example, figure 2 illustrates the differences in the daily urinary excretion of plutonium during 7 d after deep wound contamination between two physico-chemical forms: nitrate and oxide (MOX).

Graphical representation was useful to identify potential outliers. In addition, Tukey's method was used to test for outliers (Tukey 1977, Zhang 2013). As an example, at day 4 of the Pu-nitrate data from figure 2, two data points, with values of 0.21% and 0.13% percent of the administered activity respectively, were identified as outliers. However, the user has to look at the individual experimental details of these particular rats to evaluate whether these data should be removed or not from the analysis. Also other activity measurements from this animal (e.g. feces) may facilitate the decision.

2.4. Module for curve fitting of retention and excretion data

To quantify the retention and excretion kinetics, it is necessary to derive functions, usually the sum of one or more exponential terms, which give the best fit to the bioassay data (Rang *et al* 1995). Hence, the retained or excreted activity $A(t)$ in percent of the administered activity can be modelled as a function of time t (days) by the following formula:

$$A(t) = \sum_{i=1}^n a_i e^{-\lambda_i t}$$

Where n is the number of exponential terms, a_i and λ_i are the regression coefficients associated with the exponential function i . λ_i corresponds to a rate per day.

In practice, non-linear regression analysis is required to derive the regression coefficients for the mathematical function that best describes the data (Motulsky and Ransnas 1987). The computation of non-linear regression and model selection are generally challenging steps in data analysis. To facilitate these operations, specific functionalities were designed in Module #2 from figure 1. The statistical technique was selected depending on the characteristics of the data. These data consisted of repeated measurements over time on a group of subjects. The data were not normally distributed and reflected variability within and between subjects. Also, results for each subject were not independent across time. Hence, the maximum likelihood technique (Myung 2003) was implemented using mixed effects in the model (Lindstrom and Bates 1990, Tornøe *et al* 2004).

The graphical observation of data suggested that there should be one or two exponentials in the model. In addition, as the number of animals per set of experimental conditions was limited, a parsimonious model was desirable. Therefore, mono-exponential and bi-exponential functions were tested ($n = 1$ and 2 in the regression formula). The fixed effects were computed for regression coefficients (a_i and k_i) leading to coefficient estimates corresponding to the overall mean value of the response variable. The random effects (animal and time) were included to indicate that data were repeated measures on the same animal over time.

The goodness of the non-linear regression fit was assessed by different means that enable qualitative or quantitative evaluation of the closeness of the fit in relation to the data. The fitted curve superimposed on the raw data using a linear scale, and diagnostic plots of the residuals were automatically generated: standardized residuals as a function of fitted values, actual values as a function of fitted values and normal quantile-quantile plot of the standardized residuals. To complement the diagnostic plots, the akaike information criterion (AIC) was used to select the best model (Yamaoka *et al* 1978).

Figure 3 illustrates the application of the method on urinary excretion data of Am-241 in nitrate form after wound contamination from day 1 to day 7. The bi-exponential model was found to describe the data better than the monoexponential model. The biexponential model presented the smallest AIC (−43.3 versus 42.9 for the monoexponential model), which confirmed the visual comparison of the corresponding fitted curves superimposed on the data (figure 3).

2.5. Module for the implementation of biokinetic models

Biokinetic models predict the distribution of activity in the body as a function of time depending on the contamination characteristics. These compartmental models are derived from animal and available human data (Leggett 1992, Leggett *et al* 2005). The resulting predictions are the basis for the calculation of the committed effective dose (ICRP 1993, 1994, 2006, 2007). Newly generated biodistribution data are usually compared to existing biokinetic models.

To facilitate this comparison, a specific module was designed to implement the biokinetic models (Module #3 in figure 1). The set of compartments and transfer rates defining a biokinetic model can be formalized into a system of first-order differential equations. Solutions to this system are difficult to obtain analytically. Therefore, a computational integration technique was used (Soetaert *et al* 2010). This module for the implementation of biokinetic models was designed to facilitate importation of any entry model, any systemic model or any combination

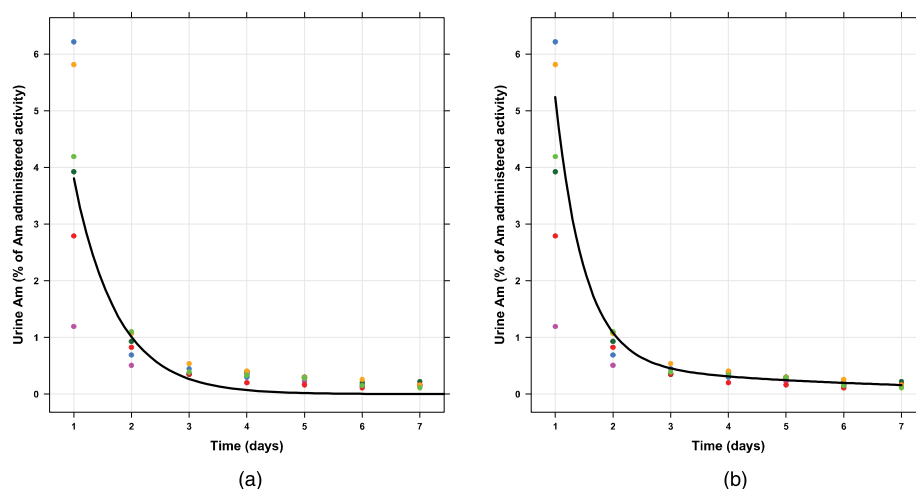


Figure 3. Comparison of the monoexponential (a) and the biexponential (b) fitted curves (black lines) of the urinary excretion data (colored points) from a wound experiment with americium in the nitrate form. Each color represents one animal.

of an entry and a systemic model. The following components were included in this module: the system of first-order equations governing the biokinetic model, the transfer rates between compartments, the initial conditions related to the activity in compartments, and the time steps for the calculation. This module was elaborated so that the user can easily change the transfer rates of an existing model or define its own model.

The NCRP wound model for soluble radiocontaminants and the ICRP systemic model of Am were implemented using this module (ICRP 1993, NCRP 2006). Transfer rate files were generated for each retention category defined in the NCRP model (weak, moderate, strong, avid) and for the systemic model. Initial conditions were set so that 100% of the activity at $t = 0$ was in the soluble compartment for the NCRP wound model and in the blood compartment for the systemic model. The predicted urinary excretion of Am from the model combination of the wound and systemic models, and from the systemic model of Am alone, were automatically represented as a function of time (figure 4) and compared with the experimental data previously introduced. Because model predictions and data were structured as *R data-frames* using the same format, plotting them on the same graph was immediate. The convenience of this comparison allowed evaluation of the category of the wound model to which the data could be associated.

2.6. Module for exploring and comparing data from the archive

The computational tool presented here provides helpful functionalities in the statistical analysis of a set of biodistribution data. Because a large number of datasets have been generated, a functionality was developed to list the available datasets and retrieve their main characteristics (Module #4 in figure 1). This feature was elaborated to get an overview of the available datasets and to foster ideas for data re-analysis, comparison, and pooling. It was designed to be applied to a set of raw data files. It extracts from each file the experimental conditions, including the range in administered activity and the number of rats per group and data point. The characteristics of the 42 datasets from the wound experiments corresponding to 42

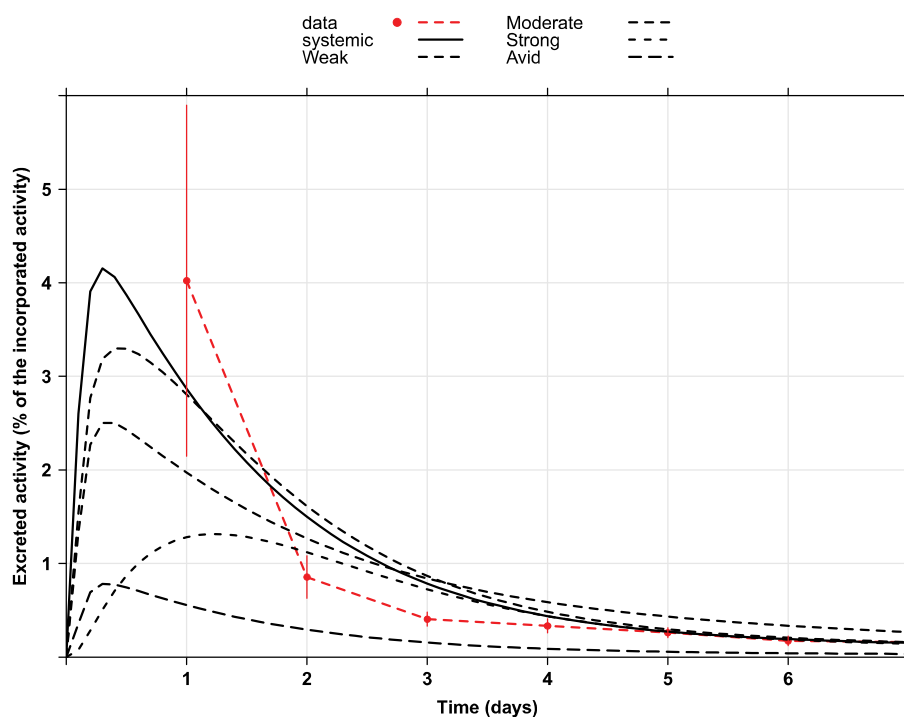


Figure 4. Comparison of biokinetic model predictions in humans with experimental rodent data for the urinary excretion of americium during 7 d after contamination. Data are from Griffiths *et al* (2012) and are represented using the group mean and standard deviation ($N = 6$). Predictions were computed using (i) the NCRP wound model for the four retention categories (Weak, Moderate, Strong, Avid) of soluble radiocontaminants (NCRP 2006), and (ii) the ICRP 67 systemic model of americium (ICRP 1993).

observations were extracted like displayed in table 3. As an example, data from experiment ‘AB_111’ consisted in daily urinary excretion up to day 7 and skeleton activity at day 7.

Data comparison is highly facilitated because all datasets share the same format as illustrated in figure 2. Also, the modules developed in this computational tool are applicable to pooled data.

3. Discussion

The analysis of biodistribution data aims at assessing the conditions that influence the biokinetics of the radiocontaminants. Contamination experiments with actinides conducted over the last decades have generated a large amount of experimental data including numerous combinations of experimental conditions. Comparison of data from diverse experimental conditions, from different laboratories, with model predictions and with human data is needed to improve the understanding and characterization of actinide biokinetics. Re-analysis of data may also be helpful to test new hypotheses (Schofield *et al* 2010). Besides, data analysis includes different levels of investigation that require increasing understanding of statistics. Therefore, it became necessary to design an approach to organize the biodistribution data and facilitate their global analysis including the functionalities listed above.

Table 3. Extract of the main characteristics from the wound experiments as contained in CSV files. One line of this table represents one set of experimental conditions.

Line #	exp_id	chemical_ form	element	act_range_ kBq	excreta_ or_tissue	time_d	number_ of_rats
1	AB_111	Nitrate	Am	4.46–6.94	Skeleton	7	6
2	AB_111	Nitrate	Am	4.46–6.94	Urine	1, 2, 3, 4, 5, 6, 7	6
3	MB_101	MOX	Am	0.0864–14.5	Skeleton	7, 30, 90	17
4	MB_101	MOX	Pu	0.0864–14.5	Skeleton	7, 30, 90	17
5	MB_101	MOX	Am	6.42–14.5	Urine	1, 2, 3, 4, 5, 6, 7, 25, 26, 30, 57, 58, 85, 86	6
6	MB_101	MOX	Pu	6.42–14.5	Urine	1, 2, 3, 4, 5, 6, 7, 25, 26, 30, 57, 58, 85, 86	6
...
42	PB_121	Nitrate	Pu	0.997–5.23	Urine	1, 2, 3, 7	24

The methodology described in this paper made the handling and analysis of data easy, fast, reproducible and traceable for the user. By including several functionalities within an integrated tool based on the *R* language, this approach avoided the use of different software packages and data formats that necessitate several manual operations (for example, ‘copy and paste’), which are prone to errors and time-consuming. Moreover, raw and transformed data, results of analysis and graphical representations were automatically saved in text, CSV, or pdf files. The reproducibility and traceability of the analyses were warranted because this methodology started directly with the raw data, and any transformation and operation conducted were done systematically through a piece of code, which prevented any manual action. Furthermore, the development of a single tool and the coding of the different operations facilitated and accelerated the data analysis that needed to be applied on numerous sets of data with different experimental conditions. Hence the use of *R* has significantly improved data analysis workflow, as compared to the use of multiple commercial software packages which would necessitate much more manual interaction.

The operations implemented were designed to conduct an analysis as comprehensive as possible. In addition to the automatic computation of group means and standard deviation, scatter plots of raw data were systematically generated (Weissgerber *et al* 2015). These representations gave an actual visualization of the data spread, which can be large with biological data. Additionally, more complex operations such as model building through non-linear regression analysis generated detailed outputs available to the user. In this work, the computational tool was applied using 42 datasets from actual wound experiments (Griffiths *et al* 2012) and the functionalities of the tool were illustrated in this paper using selected parts of these data.

The RATDOSE software package (Miller *et al* 2012) has a different objective from the present tool. Using analyses from animal data, RATDOSE was developed to assess the efficacy of chelating agents in terms of reduction in effective dose. The work in the present paper was elaborated to analyze the influence of contamination parameters on biodistribution data of radiocontaminants. It facilitates the comparison of results between different contamination conditions, laboratories and with human data. Hence, this approach is rather complementary with RATDOSE.

The computational tool described in this paper was designed to foster the creativity by facilitating the exploration of data, the re-analysis of former data, the incorporation of data from other groups, and from human cases of contamination. Hence, new hypotheses can be directly tested on previous data and remaining questions on biodistribution of actinides

after internal contamination may be elucidated from meta-analysis of these experiments. The methodology presented in this paper is therefore fully consistent with the application of the 3 Rs rule (Replacement, Reduction and Refinement) of the ethical principles applied now in experimental animal research. Furthermore, the computational tool generates a data repository, which is in line with recent efforts made worldwide to promote access to existing data such as the STORE database (Schofield *et al* 2010). However, to guarantee the security of these valuable experimental data, this repository still needs to be managed using a database system that controls any access and records any changes to the data.

Additional modules based on the same data structure could be easily included to complement the computational tool. The first application dealt with data derived from wound contamination but the approach can be easily extended to any biodistribution study of radionuclides. This extension may also include (ii) data derived from other routes of accidental contaminations, and (ii) animal or human data obtained with treatments for example to assess the efficacy of treatment protocols. In facilitating the archiving, handling and analysis of biodistribution data, this tool should contribute to the improvement of the medical response after internal contamination with radioelements.

4. Conclusion

The computational tool developed in this work facilitates and harmonizes the statistical analyses of biodistribution data from internal contamination experiments with actinides. Based on a structured organization of the data, the computational tool coded in the open-source *R* programming language implements a comprehensive analysis including several modules: descriptive analysis, retention and excretion curve fitting using non-linear regression, implementation of compartmental models and data comparison. An application of the computational tool was illustrated using data derived from wound contamination experiments with actinides. This tool was designed to make easier exploration of datasets, re-analysis of previous data, comparison across different experimental conditions and laboratories. It will also enable direct comparisons with human data, and allow conduction of meta-analyses. In addition, this methodology led to a repository essential to preserve results from animal radiotoxicological experiments. The use of this approach may be extended to broader applications related to internal contaminations with radionuclides in particular to improve the medical treatment of contaminated persons. This work is consistent with the general determination to improve the data analysis from experiments spread over time and originated from different sources.

Acknowledgment

Results presented here were obtained in the frame of the AREVA-CEA collaboration agreement.

References

- Castellani C M, Marsh J W, Hurtgen C, Blanchardon E, Berard P, Giussani A and Lopez M A 2013 *IDEAS Guidelines (Version 2) for the Estimation of Committed Doses from Incorporation Monitoring Data* (Braunschweig: EURADOS)
- Doyle-Eisele M, Weber W, Melo D R and Guilmette R A 2014 Whole-body distribution of americium in rats for different pathways of intake *Int. J. Radiat. Biol.* **90** 1068–74
- Durbin P W 2006 Actinides in animals and man *The Chemistry of the Actinide and Transactinide Elements* ed L R Morss *et al* (Dordrecht: Springer) pp 3339–440

- Freedman D 2007 *Statistics* (New York: W W Norton and Company, Inc.)
- Griffiths N M, Wilk J C, Abram M C, Renault D, Chau Q, Helfer N, Guichet C and Van der Meeren A 2012 Internal contamination by actinides after wounding: a robust rodent model for assessment of local and distant actinide retention *Health Phys.* **103** 187–94
- ICRP 1993 *Age-dependent Doses to Members of the Public from Intake of Radionuclides—Part 2 Ingestion Dose Coefficients* (Annals of the ICRP, ICRP Publication 67 vol 23) (Oxford: Pergamon)
- ICRP 1994 *Human Respiratory Tract Model for Radiological Protection* (Annals of the ICRP, ICRP Publication 66 vol 24) (Oxford: Pergamon)
- ICRP 2006 *Human Alimentary Tract Model for Radiological Protection* (Annals of the ICRP, ICRP Publication 100 vol 36) (Oxford: Elsevier)
- ICRP 2007 *The 2007 Recommendations of the International Commission on Radiological Protection* (Annals of the ICRP, ICRP Publication 103 vol 37) (Oxford: Elsevier)
- ICRP 2015 *Occupational Intakes of Radionuclides: Part 1* (Annals of the ICRP, ICRP Publication 130 vol 44) (London: SAGE)
- Leggett R W 1992 A retention-excretion model for americium in humans *Health Phys.* **62** 288–310
- Leggett R W, Eckerman K F, Khokhryakov V F, Suslova K G, Krahenbuhl M P and Miller S C 2005 Mayak worker study: an improved biokinetic model for reconstructing doses from internally deposited plutonium *Radiat. Res.* **164** 111–22
- Lindstrom M L and Bates D M 1990 Nonlinear mixed effects models for repeated measures data *Biometrics* **46** 673–87
- Métivier H, Masse R, Rateau G, Nolibé D and Lafuma J 1989 New data on the toxicity and translocation of inhaled $^{239}\text{PuO}_2$ in Baboons *Radiat. Prot. Dosim.* **26** 167–72
- Miller G, Bertelli L, Klare K, Weber W, Doyle-Eisele M and Guilmette R 2012 Software for empirical building of biokinetic models for normal and decorporation-affected data *Health Phys.* **103** 484–94
- Motulsky H J and Ransnas L A 1987 Fitting curves to data using nonlinear regression: a practical and nonmathematical review *FASEB J.* **1** 365–74
- Myung I J 2003 Tutorial on maximum likelihood estimation *J. Math. Psychol.* **47** 90–100
- NCRP 2006 Development of a biokinetic model for radionuclide-contaminated wounds and procedures for their assessment, dosimetry, and treatment: recommendations of the National Council on Radiation Protection and Measurements *NCRP Report No. 156*
- Priest N D 2001 Toxicity of depleted uranium *Lancet* **357** 244–6
- R Core Team 2015 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing (Vienna) www.R-project.org/
- Rang H, Dale M M, Ritter J and Moore P 1995 *Pharmacology* (London: Churchill Livingstone)
- RStudio Team 2015 *RStudio: Integrated Development for R* (Boston, MA: RStudio, Inc.) www.rstudio.com/
- Sarkar D 2008 *Lattice: Multivariate Data Visualization with R* (New York: Springer)
- Sarkar D 2015 *Package 'Lattice'. Version 0.20-33*
- Schofield P N, Tapio S and Grosche B 2010 Archiving lessons from radiobiology *Nature* **468** 634
- Soetaert K, Petzoldt T and Setzer R W 2010 *Solving Differential Equations in R: Package deSolve J. Stat. Software* **33** 25
- Tornøe C W, Agerso H, Nielsen H A, Madsen H and Jonsson E N 2004 Pharmacokinetic/pharmacodynamic modelling of GnRH antagonist degarelix: a comparison of the non-linear mixed-effects programs NONMEM and NLME *J. Pharmacokinet. Pharmacodyn.* **31** 441–61
- Tukey J 1977 *Exploratory Data Analysis* (Reading, MA: Addison-Wesley)
- Van der Meeren A, Gremy O, Renault D, Miroux A, Bruel S, Griffiths N and Tourdes F 2012 Plutonium behavior after pulmonary administration according to solubility properties, and consequences on alveolar macrophage activation *J. Radiat. Res.* **53** 184–94
- Van der Meeren A, Tourdes F, Gremy O, Grillon G, Abram M C, Poncy J L and Griffiths N 2008 Activation of alveolar macrophages after plutonium oxide inhalation in rats: involvement in the early inflammatory response *Radiat. Res.* **170** 591–603
- Weber W, Doyle-Eisele M, Melo D R and Guilmette R A 2014 Whole-body distribution of plutonium in rats for different routes of exposure *Int. J. Radiat. Biol.* **90** 1011–8
- Weissgerber T L, Milic N M, Winham S J and Garovic V D 2015 Beyond bar and line graphs: time for a new data presentation paradigm *PLoS Biol.* **13** e1002128
- Yamaoka K, Nakagawa T and Uno T 1978 Application of Akaike's information criterion (AIC) in the evaluation of linear pharmacokinetic equations *J. Pharmacokinet. Biopharm.* **6** 165–75
- Zhang Z 2013 Advancements of outlier detection: a survey *ICST Trans. Scalable Inf. Syst.* **13** e2