



Exposure Factors vs. Bioaccessibility in the Soil-and-Dust Ingestion Pathway: A Comparative Assessment of Uncertainties Using MC2D Simulations in an Arsenic Exposure Scenario

F. Barrio-Parra¹ · H. Serrano García¹ · M. Izquierdo-Díaz¹ · E. De Miguel¹

Received: 21 July 2022 / Revised: 17 November 2022 / Accepted: 22 December 2022 / Published online: 31 January 2023
© The Author(s) 2023

Abstract

Human Health Risk Assessment (HHRA) is a widely applied method to make decisions about the environmental status of sites affected by toxic substances. Its conclusions are affected by the variability and uncertainty of the input variables in the HHRA model. The aim of this work is to apply an algorithm based on 2D Monte Carlo simulations to integrate the variability and uncertainty of exposure factors, concentration, and bioaccessibility, reported by various information sources, to assess and compare their influence on the risk outcome. The method is applied to a specific case study of exposure of children to arsenic from accidental soil ingestion in a residential setting in the city of Madrid (Spain) by combining information from 12 studies. The consideration of the variability and uncertainty of the exposure parameters in the Baseline Risk Assessment (BRA, deterministic) resulted in a greater reduction in the numerical value of risk estimations than that produced by considering only the bioaccessibility factor. The results of the Probabilistic Risk Assessment (PRA) showed that the risk distribution was more sensitive to the variabilities of the accidental soil intake rate and the total arsenic concentration than to other variables such as bioaccessibility. In this case study, the uncertainty introduced by using the "default" reasonable maximum exposure factors in the HHRA model and the variability of the concentration term produce overestimates of risk that are at least in the range of those produced by omitting the bioaccessibility term. Thus, the inclusion of bioaccessibility is, alone, insufficient to improve the HHRA since the selection of the exposure factors can significantly affect the estimates of risk for the soil ingestion pathway. In other sites or for other contaminants, however, the role of the uncertainties associated with the bioaccessible fraction could be more pronounced. The method applied in this work may be useful in updating exposure factors to reduce uncertainties in HHRAs.

Keywords Human Health · Probabilistic Risk Assessment (PRA) · Urban Scenario · Arsenic

Introduction

The Human Health Risk Assessment (HHRA) methodology is a widely established process for making decisions on the potential adverse health effects that may arise in the population exposed to chemicals in different environmental compartments (i.e., soil, surface water, groundwater and air) (USEPA 1989). The point estimate of risk using deterministic models (Deterministic Risk Analysis or DRA) is the first tier in the tiered approach presented by USEPA (2001)

and is the procedure most commonly employed by risk analysts. These analyses recurrently use reasonable maximum exposure (RME) parameters for the estimation of the baseline risk, which is understood as a conservative estimate of the risk of the studied population (USEPA 2011a, 2014a). Sometimes, these exposure parameters may not necessarily be representative of the study population, especially when scenarios or local receptors show particular activities (Barrio-Parra et al. 2019; De Miguel et al. 2017; Dourson et al. 2013), so the worst-case assumption introduces an unknown degree of uncertainty in risk assessments (Jager et al. 2001). The next levels in the tiered approach allow analysing the role of variability and uncertainty of the exposure variables in the uncertainty of the risk model output through Probabilistic Risk Assessment (PRA). This methodology allows characterising confidence in risk estimates and identifying

✉ F. Barrio-Parra
fernando.barrío@upm.es

¹ Prospecting & Environment Laboratory (PROMEDIAM),
Universidad Politécnica de Madrid, 28003 Madrid, Spain

the key parameters that determine their distribution in the population. This information can be very useful to risk managers in decision making. (Hammonds et al. 1994; ODEQ (Oregon Department of Environmental Quality) 1999; USEPA 1997)).

In addition to the exposure parameters involved in HHRA, the mode of determination of the concentration term of the toxic substance in the exposure medium also introduces uncertainty in the risk assessment. This uncertainty is associated with the sample size, the variability of the concentration data at the site, and its bioaccessibility. Commonly, the uncertainty associated with the variability of the concentration data and the sample size is taken into account through the calculation of the UCL95, understood as the upper limit of the mean at a 95% confidence level (USEPA 1992). Some freely available software tools facilitate the calculation of this uncertainty in risk output, such as the ProUCL tool (USEPA 2013). Oral bioaccessibility is defined as the fraction of pollutant that is extracted (put into solution) by digestive fluids in the salivary system and the gastrointestinal tract and it differs from the concept of bioavailability in that the latter refers to the fraction of ingested compounds that is absorbed and reaches the systemic circulation (Denys et al. 2009; USEPA 2012). The consideration of bioaccessibility reduces the overestimation of risk in scenarios where the main route of entry into the organism is accidental ingestion of soil and dust (Guney et al. 2013; Hamad et al. 2014; Li et al. 2018; Martínez-Sánchez et al. 2013), which is applicable to those substances whose toxicity data are derived from toxicological studies in which the substance is delivered in dissolved form (Izquierdo-Díaz et al. 2015; Ruby et al. 1999). This is why the incorporation of the term bioaccessibility into risk analysis has gained interest in recent years (Ge et al. 2002). The literature describes several analytical methods for the estimation of this parameter (e.g. Simplified Bioaccessibility Extraction Test (SBET) (Ruby et al. 1999), Physiologically Based Extraction Test (PBET) (Oomen et al. 2003) also known as UBM (Unified BARGE Method) (Wragg et al. 2009) or Hydrochloric acid-extractable fraction (Rasmussen et al. 2008)). The term bioaccessibility will be associated with uncertainties arising from the choice of the analytical method for its determination (accuracy of bioaccessibility determination) and its precision.

Inclusion of the bioaccessibility parameter in the risk equations can significantly reduce its magnitude. However, there are other sources of uncertainty that can significantly affect the indeterminacy of risk, such as those associated with the toxicity value, environmental background exposure (Barrio-Parra et al. 2018; Li and Li 2021; Stanfield et al. 2021), or, as already mentioned, those associated with the exposure parameters (Mingot et al. 2011). Regarding the latter, the uncertainty produced by the use of default values can be reduced through their characterisation by local studies

(e.g. surveys on exposure frequency, consumption of vegetables grown at the site, or body weight (Barrio-Parra et al. 2019; Cao et al. 2016; De Miguel et al. 2007; Iribarren et al. 2009; Jiménez-Oyola et al. 2021)). In some cases it may be interesting to characterise these variables through previous studies due to the high robustness of the database on a specific population (e.g., studies to characterise life expectancy at birth or on the distribution of body weights among different age groups in a population (i.e., Burmaster and Crouch 1997)). It may also be interesting to use literature information on the characterisation of exposure variables that requires data collection methods that are more difficult to implement in risk assessment studies. An example of such a parameter is the accidental soil and dust ingestion rate. This exposure factor can be characterised through the joint study of information collected by different papers using different methodologies (i.e., tracer element, biokinetic model comparison, or activity pattern methods (USEPA 2011a)). Given that there are different sources of variability and uncertainty associated with exposure, concentration, and bioaccessibility parameters, a comparative study of the impact of each one on the indeterminacy of risk is of interest.

Recent studies have used Monte Carlo Simulation (MCS)-based methods to investigate the impact of variability of exposure factors and/or total concentrations and bioaccessibility on risk estimates (Hamad et al. 2014; Liang et al. 2013; Rajasekhar et al. 2018; Sadeghfam et al. 2021; Wang et al. 2020). However, there is a lack of applications of MCS to study the impact of variability and uncertainty of input variables on the risk model when reported by different sources of information. For example, integration of information concerning the variability of (i) the concentration term of a trace element from different studies (local and regional), (ii) the bioaccessibility in the study area estimated by different methodologies, and (iii) the exposure factors characterised by different studies. Furthermore, the probabilistic modelling of each variable in the risk model should reflect the uncertainty associated with the statistical power of the information source (sample size). Therefore, in order to analyse and compare the sensitivity of risk outcomes to concentration terms, bioaccessibility, and exposure factors, it seems necessary to apply an MCS-based method that allows the construction of data distributions considering the variability and uncertainty reported by different data sources. In this way, the impact on risk estimation of variability and uncertainty of parameters obtained from local information, as well as those characterized from information from different studies available in the scientific literature, could be compared.

For all of the above reasons, the aim of this work is to apply an algorithm based on 2D Monte Carlo simulations to assess and compare the impact of the variability and uncertainty of the exposure factors and the concentration and bioaccessibility terms reported by different sources

of information in quantitative risk assessments, as well as its application to an illustrative case example involving the exposure of children to arsenic in soil of municipal playgrounds.

Background: 2D Monte Carlo Simulations

The method applied in this work pursues the analysis and comparison of the uncertainties introduced in the exposure assessment phase of the risk characterisation (Dong et al. 2015; International Programme on Chemical Safety, 2008) according to the scenario outlined in “[Study area and definition of the risk model](#)”. Specifically, the effect of variability and uncertainty on all exposure variables (except exposure duration, ED in years) is assessed. In the case of bioaccessibility (BA, unitless), the variability and uncertainty associated with its determination through the three analytical methodologies (SBET, HCl, and PBET extractions) are also assessed. According to the recommendations of USEPA (2001), variability and uncertainty should only be considered for exposure assessment, with SF (oral slope factor) and RfD (Reference Dose) as point values. Variability can be defined as the true heterogeneity or diversity that characterizes an exposure variable or response in a population. The concept of uncertainty is related to the lack of knowledge about a specific variable or parameter (e.g., the mean value of the body weight in a population). Increasing the quality and quantity of the information about a given variable (e.g., increasing the sample size) will increase the confidence on the value of the parameters that describe the variability of that variable. Increasing sample size will not reduce variability, but it will reduce the uncertainty about the parameters that define it (USEPA 2001).

The influence of exposure factors on the uncertainty of the risk outcome will be estimated through a Probabilistic Risk Assessment (PRA) (ODEQ, 1999; USEPA 1997). The PRA proposed in this work is based on Monte Carlo Analysis (MC) whereby many random numbers (i.e. law of large numbers, in this study $N_i = 10^5$) are generated to represent the variability of exposure factors, the concentration term, and the bioaccessibility term. These random values are entered into the exposure equations to generate risk distributions (Frey and Rhodes 1996; Harris and Horn 1996). The uncertainty of the outcome is represented by the variability of the risk distribution produced by the collective variation of the input variables. Therefore, the uncertainty of the risk (understood as its variability within the population studied) can be analyzed using histograms, empirical cumulative distribution function (CDF) plots, and descriptive statistics (Millard 2013). The sensitivity of the results to the variability of the input terms to the risk model can be studied through correlation analysis (Iman and Helton 1988). Once sensitive variables have been identified, their significance is

assessed, understood as the contribution of their variability to uncertainty in the assessment of risk outcomes (Hamby 1994).

In the PRA, knowing the parametric function that represents the variability of each exposure factor (Probability Distribution Function, PDF), vectors of values can be generated through the application of functions in R (R Core Team 2022) such as those available in the “truncdist” library (Novomestky and Nadarajah 2016). This study applies 2-Dimensional Monte Carlo Analysis (2D-MC) as a method of generating random data vectors that introduces uncertainty in the parameters that define the PDFs of the exposure variables (Frey and Rhodes 1996; Pouillot and Delignette-Muller 2010; Sadeghfam et al. 2021). Previous studies have applied the 2D-MC method to perform exposure simulations in different scenarios (i.e., Jang et al. 2009; Simon 1999; Vadali et al. 2009; Wu et al. 2019). To the best of the authors’ knowledge, there are no applications of the 2D-MC method for the simulation of exposure scenarios that consider variability and uncertainty from various sources of prior information. Thus, this work seeks to apply the 2D-MC simulation to HHRA to integrate the variability and uncertainty collected from different sources of information (local studies and others of regional or international scope). In 2D-MC, one or more parameters of the PDF are not fixed but are defined by another specific distribution (Millard 2013). Uncertainty is represented through the random generation of parameters that define a PDF, for example, the parameter “mean” (μ) in a normal distribution. That is, a new random number generation function is introduced to obtain a random value of μ in a first MC step. This random value of μ will be used on a second level of MC to obtain a random number for the exposure variable in question. The process is repeated as many times as random numbers are generated, thus jointly representing the variability and uncertainty of a given exposure factor/ total concentration/ bioaccessibility.

“[Correlation analysis of input variables](#)” describes the methods applied to generate random numbers and to analyze the sensitivity and significance of the exposure factors to the risk scenario. These methods are implemented within the R framework (R Core Team 2022).

Materials and Methods

Study Area and Definition of the Risk Model

An exposure scenario involving the exposure of children (aged 6 months to 6 years) to arsenic in urban playground soil was considered to test the applicability of the proposed methodology. The starting data come from studies carried out by Mingot et al. (2011) and De Miguel et al. (2012) for the city of Madrid (Spain), focusing on the characterisation of

the risk of exposure to Arsenic (As). Arsenic is considered for its potential to cause numerous health disorders (Das et al. 2013), including cancer and circulatory disorders (Dudka and Miller 1999) and occupies the first position in the list of priority harmful pollutants for human health (ATSDR, 2019).

Children's exposure to trace elements in playgrounds can occur through three main routes: (i) direct ingestion of soil particles; (ii) inhalation of suspended particles; and (iii) dermal absorption in particles adhering to exposed skin (De Miguel et al. 2007). In the present work, accidental ingestion of soil particles has been prioritised as the most relevant route of exposure, considering the tendency of children to play in contact with the soil and to place objects and their hands in their mouths. Several studies have shown that ingestion is the route that contributes the most to the estimation of risk in a recreational setting, while the influence of dust inhalation and dermal contact can be considered insignificant (De Miguel et al. 2012, 2007; Dudka and Miller 1999; Han et al. 2020; Izquierdo-Díaz et al. 2015; Penteado et al. 2021). The dose received through soil intake (I_s) is mathematically calculated with the following expression:

$$I_s = \frac{CS * BA * IR * CF * EF * ED}{BW * AT}, \quad (1)$$

where CS is the concentration of the contaminant in soil ($\text{mg}\cdot\text{kg}^{-1}$), BA the bioaccessibility of the contaminant (–) IR the soil intake rate ($\text{mg}\cdot\text{day}^{-1}$), CF the conversion factor ($10^{-6} \text{ kg}\cdot\text{mg}^{-1}$), EF the exposure frequency ($\text{days}\cdot\text{year}^{-1}$), ED the exposure duration (year), BW the body weight (kg) and AT the average time (days). The value of AT varies depending on the toxic effect being assessed. For non-carcinogenic effects, exposure is averaged over the duration of exposure. For carcinogenic effects, exposure is averaged by life expectancy (USEPA 1989). To obtain the Hazard Quotient (HQ or non-carcinogenic risk) for oral exposure to a toxicant, the dose calculated with Eq. 1 is divided by its reference dose (RfD, $\text{mg}\cdot\text{kg}^{-1}\cdot\text{day}^{-1}$) (Eq. 2), while, to quantify the level of carcinogenic risk, the dose received is multiplied by its slope factor (SF, $\text{mg}^{-1}\cdot\text{kg}\cdot\text{day}$) (Eq. 3). The RfD and SF values for arsenic are specified as $3\cdot 10^{-4} \text{ mg}\cdot\text{kg}^{-1} \text{ day}^{-1}$ and $1.5 \text{ mg}^{-1}\cdot\text{kg}\cdot\text{day}$, respectively (USEPA 2022).

$$HQ = \frac{I_s}{RfD} \quad (2)$$

$$Risk = I_s \times SF \quad (3)$$

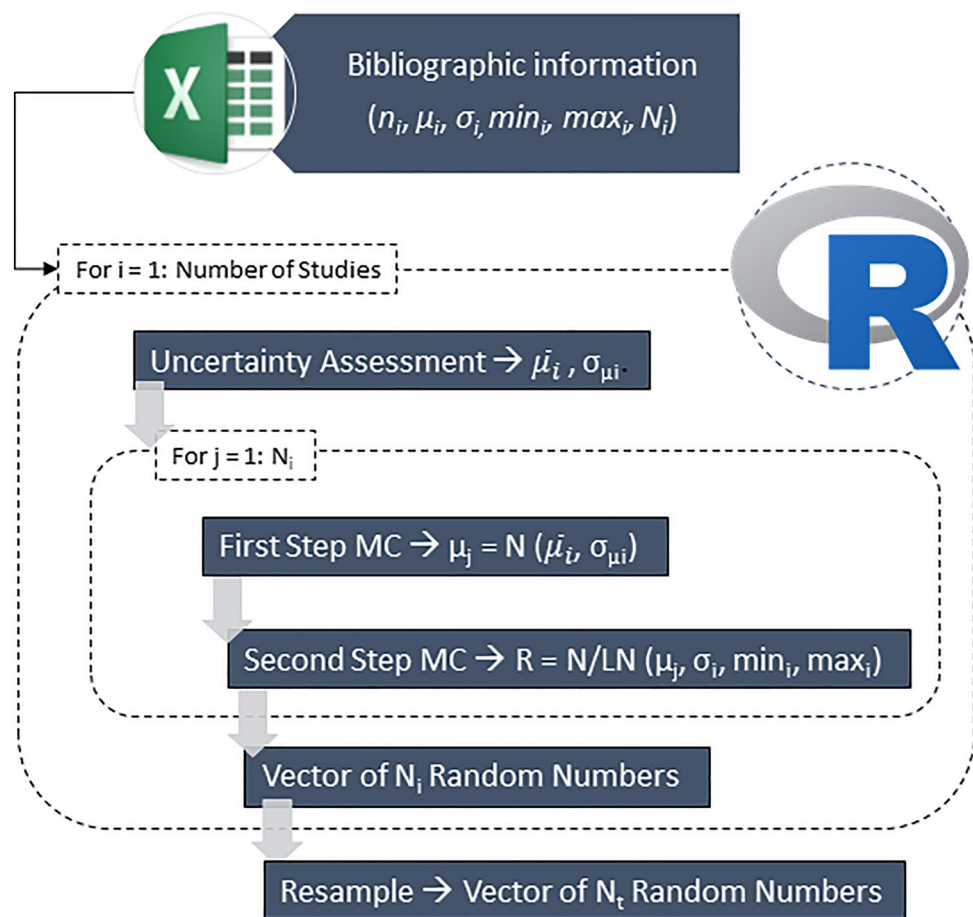
Modelling Parametric Uncertainty and Variability Through 2D Monte Carlo Analysis

The proposed algorithm for the generation of data vectors integrating variability and uncertainty of exposure factors

based on information from different papers is depicted in Fig. 1 and is organised in the following steps:

1. Reading information: The routine begins with the reading of an Excel file containing in rows the information from each bibliographic source on a given variable, including in its columns information on the sample size (n_i), the mean (μ_i), standard deviation (σ_i) and the minimum (min_i) and maximum (max_i) values obtained in the study. A column is also included with the number of random data to be generated for each study (N_i). The total number of random numbers to be generated is fixed beforehand (in this study $N_i = 10^5$), so the number of data to be generated for each study (N_i) is obtained proportionally according to the number of bibliographic references considered in each variable and it must be fulfilled that $N_i = \sum N_i = 10^5$.
2. Characterisation of the uncertainty in each study: To characterise the uncertainty in the determination of the parameters that define the PDF reported by a bibliographic source, this work applies functions that, from a Bootstrap approach, perform 10^4 iterations in which n_i random numbers are generated following the PDF defined by the bibliographic source (normal and log-normal distributions have been considered in this work). In each iteration, the average of the generated random n_i numbers is obtained, obtaining at the end of the loop a vector of 10^4 values of μ . The PDF of this vector (assumed to be normal and therefore defined by a mean value ($\bar{\mu}_i$) and standard deviation ($\sigma_{\mu i}$) characterises the uncertainty in the determination of the mean in the exposure variable, total concentration or bioaccessibility in a literature source. Therefore, for each study (row in the Excel file, i), the function is run with the resulting parameters $\bar{\mu}_i$ y $\sigma_{\mu i}$.
3. Execution of the 2D Monte Carlo Simulation: Monte Carlo simulation is performed in two stages for each study, repeating the process from $j=1$ to $j=N_i$.
 - a. First level of Monte Carlo Simulation: Once the parameters representative of the uncertainty in the determination of the mean ($\bar{\mu}_i$ y $\sigma_{\mu i}$) have been estimated, a random number μ_j ($N(\bar{\mu}_i, \sigma_{\mu i})$) is generated by applying the *rnorm* function of the "stats" library. (R Core Team 2022).
 - b. Second-level Monte Carlo simulation. From μ_j and σ_i a random number R is obtained following a Normal or Log-normal distribution (depending on the PDF specified by the bibliographic source evaluated) applying the *rtrunc* function of the "truncdist" library (Novomestky and Nadarajah 2016) considering as truncation limits the fifth and 95th percentiles reported in each study (if these values are absent,

Fig. 1 Workflow for 2D Monte Carlo simulation



the minimum and maximum values are used) thus avoiding the possibility of generating impossible values (i.e., negative) or outside the limits of each study. (Harris and Horn 1996).

4. Permutation of data: Once the N_i random values have been obtained, they are permuted by executing the *sample* function without replacement.

Correlation Analysis of Input Variables

In the PRA, to ensure that the values of the input variables are independent and that random sampling of the data can therefore be done independently, it is necessary to perform a correlation analysis (Hammonds et al. 1994; USEPA 1997). This correlation can be assessed by calculating the Spearman rank correlation coefficient with the *cor.sign* function of the "StatDa" library (Filzmoser 2015). If the obtained correlation is below 0.6, it is assumed that the relationship between the variables is weak and, therefore, random sampling can be performed independently (Barrio-Parra et al. 2019; Burmaster and Anderson 1994; Harris and Horn 1996). In this work, exposure parameters (i.e., IR, BW, AT, EF, see "Local data sources") are characterized by PDFs

from different bibliographic sources. Thus, it is not possible to perform a correlation analysis, and the independence of these variables is assumed.

Spearman's analysis yielded correlation values below 0.1 between total As concentration and bioaccessibility percentages determined by SBET, PBET and HCl methods by Mingot et al. (2011) (in all cases, p values denoted low statistical significance). Consistent with previous works (De Miguel et al. 2012; Li et al. 2018; Mingot et al. 2011), the hypothesis of independence between total As concentration and percentage of bioaccessibility is validated. Therefore, the percentage of bioaccessibility is solely dependent on the edaphic properties of the soil. The independence of these variables allows the use of independent and complementary sources of information on the geochemical variability of arsenic in soils with similar edaphic properties.

Local Data Sources

Annex A summarises the bibliographical information used for the generation of random variables. In general, sources of information where the PDFs of the input variables are specified have been selected, with an emphasis on those studies of a local nature. Specifically, the terms concentration

(CS), bioaccessibility (BA) and exposure frequency (EF) have been able to draw on data from the aforementioned works by Mingot et al. (2011) and De Miguel et al. (2012), as well as others carried out in the Community of Madrid (De Miguel et al. 2002, 2007).

Total As concentrations in the soil of playgrounds in the city of Madrid have been previously described and discussed by Mingot et al. (2011). The values obtained by these authors did not show significant differences when performing a Tukey test for a comparison of means at 95% confidence (p -value = 0.65) with the concentrations in soils formed from the arkose bedrock on which the city sits (De Miguel et al. 2002) (Annex A). As no differences were found in the means of the two studies, it was assumed that the data population from the playgrounds belongs to the same lithological unit as the lithological unit in which the playgrounds are located, thus allowing the information from both sources to be included in the study of the variability of As concentration in the study area.

This work takes into account the bioaccessibilities of arsenic evaluated by Mingot et al. (2011) of soils with edaphic characteristics similar to those of urban parks in the city of Madrid, based on three different methodologies: two gastric (SBET and extraction by HCl) and one gastrointestinal (PBET). Although As released in the simulated gastric medium reaches average bioaccessibilities between 63% (SBET) and 52% (HCl), a fraction of As is recaptured and removed from the solution under intestinal conditions, reducing the effective absorption of As to 42% (PBET). Furthermore, the authors determined significant positive correlations between the three methodologies ($r > 0.77$), and especially between the results of the PBET protocol and those of the HCl extraction ($r = 0.893$).

To select the PDF that best fits each data series (total arsenic concentration and bioaccessibility percentages), the `gofTest` function of the `EnvStats` library (Millard 2013) is employed. The results of the goodness-of-fit tests (Shapiro–Wilk test) yielded statistically significant results for log-normal distributions in the case of the total concentration data series and normal distributions for the bioaccessibility percentages. The log-normal distribution obtained for the fit of the playground data is consistent with that obtained for the regional geological unit (De Miguel et al. 2002). The representation of the variability of the bioaccessibility percentage through a normal distribution indicates that it is produced by random (unbiased) errors in the analytical determinations (Harris and Horn 1996) and is therefore a representation of the experimental indeterminacy produced by the application of each laboratory method.

The frequency of exposure (FE) of children in urban playgrounds was estimated from a survey of 36 families by De Miguel et al. (2007) and meteorological data from the Community of Madrid. According to the questionnaire

results, children up to 6 years of age visit a playground on average 4 days a week. The number of weeks per year that a child visits these playgrounds was estimated from the average number of rain-free days in the municipality of Madrid for the period 2007–2020 (280 days or 40 weeks), minus 4 weeks that were considered as a holiday period away from the usual residence, giving a total exposure frequency of 136 days-year⁻¹.

Bibliographic Data Sources

The information used to characterise the rest of the exposure variables comes from bibliographic data sources (Annex A). Regarding body weight (BW), Spanish child studies have been taken into account considering age ranges from 6 months to 6 years. The age interval was selected to match the periods studied in the characterisation of the accidental rate of soil and dust. Each study includes information corresponding to different age ranges and gender (Carrascosa et al. 2008a; 2008b; Durá Travé, 2012). For each age range studied, the lowest and highest reported percentiles have been established as lower and upper truncation limits (2nd and 98th percentiles in Carrascosa et al. (2008a; 2008b) and the 3rd and 98th percentiles in Durá Travé (2012)).

For the soil intake rate (IR), we have established as a criterion the inclusion of studies in which the distribution of the data is indicated independently of the population studied (no reports on the soil intake rate have been found in the study area). Most of the studies evaluated are compiled in USEPA (2017) and other reviews (Moya and Phillips 2014). Among them, some have been discarded (Calabrese et al. 1991; Davis et al. 1990; Stanek & Calabrese 1995) due to containing redundant information and/or information compiled in other more general studies (Stanek et al. 2012). Likewise, other sources that considered anomalous behaviours such as pica or geophagy have been discarded (i.e. Calabrese et al. 1991).

Regarding the duration of exposure (ED), a single value of 5.5 years has been considered, corresponding to the maximum age of the children defined for the exposure scenario minus the 0.5 years of the period in which there is no information on the rate of accidental soil ingestion.

Finally, the averaging time (AT) for carcinogenic substances has been obtained through a study of the decomposition of differences in life expectancy at birth according to the differential mortality of the entire Spanish population in 2009 (Goerlich Gisbert 2012).

Baseline Risk Assessment (BRA)

To obtain comparative values of the modelling results of the exposure parameters by 2D Monte Carlo simulations, as well as to compare the results of the risk uncertainties obtained

with the PRA, a baseline risk assessment (BRA) is initially performed in the scenario (Barrio-Parra et al. 2019; Harris and Horn 1996). BRA considers the total concentration and the bioaccessibility percentages calculated by the three analytical methods in the study area (Mingot et al. 2011). These risk assessments should be considered as estimates of the upper limits of the risk to which the population is exposed (Hammonds et al. 1994), and therefore allow the upper limits obtained through the PRA to be contrasted. The concentration term was calculated as the 95% upper confidence limit (UCL95) as a conservative estimator of the mean concentration to which the child population visiting the playgrounds is exposed. This estimate was made parametrically assuming a log-normal distribution of the data (see “Local data sources”) (USEPA 1992). The mean bioaccessibility percentages are those obtained through linear regression models between total arsenic concentrations and bioaccessible concentrations for each of the applied methods. For the rest of the parameters, the default values recommended by the USEPA (2014a) have been considered, which correspond to values representative of the maximum reasonable exposure (Annex B).

Assessment of Uncertainty, Sensitivity, and Important Variables in the PRA

Once the vectors of 10^5 random data for each variable have been generated according to the procedure described in “Modelling parametric uncertainty and variability through 2D Monte Carlo analysis”, they are introduced into Eq. 1 to calculate the dose received by children playing in urban playgrounds through accidental ingestion of arsenic in soil and dust ($\text{mg}\cdot\text{kg}^{-1}\cdot\text{day}^{-1}$), averaged for systemic and carcinogenic effects. Through this distribution of the estimated dose and by means of the toxicological data and Eqs. 2 and 3, the distributions of hazard quotients and carcinogenic risk are obtained for the studied population.

The uncertainty in the risk estimates is represented by its variability and is studied through their graphical representation by cumulative distribution function graphs (CDFs) and the obtaining of descriptive statistics (minimum, maximum, range, median, 95th percentile) as well as the percentage of data that exceed the admissible risk threshold in each case ($\text{HQ} > 1$ in the case of systemic risk and carcinogenic risk greater than 10^{-5}). To facilitate the discussion of the role of bioaccessibility in risk indetermination, two groups of simulations are proposed: on the one hand simulations in which only the total concentration is considered and on the other hand, the total concentration and the percentage of bioaccessibility determined by different methods.

Sensitive variables are those that have a substantial influence on the risk distribution (Hamby 1994; Millard 2013). To perform a quantitative analysis of the sensitivity of the

results to variability and uncertainty of the input variables, and thus identify priority uncertainties (Iman and Helton 1988; USEPA 2014b), Spearman correlation coefficients between model inputs and outputs are calculated. The model has a linear response to all variables, so their influence on the outcome will be determined solely by the shape of their PDF (kurtosis and skewness). Important variables are those whose amount of uncertainty or variability contributes substantially to risk variability (Hamby 1994; Millard 2013). Once the most sensitive variables have been identified, the impact of the range of their variability in the percentage of the population on the unacceptable risk thresholds is assessed (importance analysis), including a comparative analysis of the variability of the exposure factors with the proposed default values for reasonable maximum exposure scenarios.

Results and Discussion

Baseline Risk Assessment

Table 1 shows the results of the deterministic risk analysis estimated from reasonable maximum exposure parameters for the oral exposure scenario in playgrounds considering the total concentration (UCL95) and its correction considering the bioaccessibility obtained with different methods (HCl extraction, SBET and PBET). In no case does the hazard quotient (HQ) exceed the acceptability threshold. However, for carcinogenic risk, the acceptability threshold (set at 10^{-5} by most regulatory agencies) is exceeded if the total concentration term is entered without considering the percentage of bioaccessibility. On average, the risk term (systemic and carcinogenic) is halved when the bioaccessibility of arsenic is included in the scenario.

In other works in which arsenic bioaccessibility in playgrounds has been assessed, the risk could be reduced by 70%–90%, since bioaccessibility is a term highly dependent on soil properties (De Miguel et al. 2019, 2012; Glorenec et al. 2012; Ljung et al. 2007). In studies carried out in other scenarios in the city of Madrid (Izquierdo-Díaz et al. 2015), the consideration or not of the bioaccessibility term was decisive in the acceptability/unacceptability of the risk.

Table 1 Baseline Risk Assessment results obtained with total and bioaccessible Arsenic concentrations

As concentration	HQ	Risk
Total	0.31	$1.19\cdot 10^{-5}$
HCl	0.16	$6.21\cdot 10^{-6}$
SBET	0.19	$7.53\cdot 10^{-6}$
PBET	0.13	$5.02\cdot 10^{-6}$

Although the influence of bioaccessibility on the risk outcome may be decisive, it is still interesting to contextualize its influence against the variability and uncertainty of the other input variables in the model (Mingot et al. 2011), as well as the differences arising from the use of variables characterized from local studies versus reasonable maximum exposure parameters. Given that the estimates of carcinogenic risk in the scenario evaluated are closer to the limits of unacceptability than those of non-carcinogenic risk, the PRA is considered only for carcinogenic risk.

Effect of Sample Size on the Variability of the Mean

The random number generation method presented in this paper assumes that the values of the mean, obtained by the generation of random vectors of length equal to the sample size reported by each study, follow a normal distribution. Figure 2 shows three examples that allow us to validate this hypothesis. Here, one can see the result of a random value simulation in three studies of soil intake rates (studies #3 (Özkaynak et al., 2011), #5 (Thompson and Burmaster, 1991) and #7 (Davis and Mirick (2006) in (Annex A) with different sample sizes (N , left column) and the result of the mean distributions obtained after 10,000 simulations (right column). The histograms in the right column show how the variability of the mean follows a presumably normal distribution, centred on the mean value reported by the study and with a standard deviation inversely proportional to the sample size. This reflects how the proposed methodology introduces the uncertainty of the survey (inversely proportional to the sample size) into the process of generating random data.

Random Variable Generation

Figure 3 shows the histograms of the 10^5 random data generated following the proposed 2D Monte Carlo-based variability and uncertainty modelling methodology for the exposure parameters considered. Annex C includes the descriptive statistics of these distributions.

The resulting distribution of accidental soil and dust ingestion rate (IR) is very asymmetric and has marked tails, a consequence of the use of log-normal distributions reported in the literature. In the distribution, it has been possible to generate data above the value proposed by the USEPA (2011a), with the probability of exceeding the value of 200 mg day^{-1} being 0.15%. The 95th percentile of the distribution obtained in this study ($86 \text{ mg}\cdot\text{day}^{-1}$) lies well below the reasonable maximum exposure value of $200 \text{ mg}\cdot\text{day}^{-1}$ (Annex B), and is closer to the recommendations for the assessment of the central tendency ($84 \text{ mg}\cdot\text{day}^{-1}$) according to USEPA (2017) and with the 95th percentile estimated by (Stanek et al. 2012) in a study of eight tracer elements ($79 \text{ mg}\cdot\text{day}^{-1}$). These results reflect the variability between

studies and the degree of indeterminacy of the high percentiles of IR (the range of the plausible values spans 3 orders of magnitude, Annex C). The use of the 95th percentile of the data distribution obtained by combining information from different studies would result in a reduction of approximately 60% of the risk in a deterministic analysis compared to the proposed value for the reasonable maximum exposure.

The values generated to represent the exposure frequency (EF) have been estimated with a log-normal distribution obtained from a survey of the exposed population (De Miguel et al. 2007). It is very remarkable that the maximum values obtained ($252 \text{ days}\cdot\text{year}^{-1}$) are below the reasonable maximum exposure values proposed by USEPA (2014a, b) for a residential scenario ($350 \text{ days}\cdot\text{year}^{-1}$). According to the USEPA (1991) definition, residential scenario conditions can be assumed as long as there are occupied residences adjacent to the study site, with residents being frequently exposed to the contaminated environment. These assumptions consider long-term daily exposure and produce the highest potential risks. In our case study, we have playgrounds in urban areas, so the conceptual model of the scenario is fairly close to the definition of the residential scenario. Despite the proximity of the contact zone to residences, playgrounds follow a specific pattern of use that must be characterised in each case. In the case of assuming the residential scenario with the maximum reasonable EF, we would introduce an overestimation of risk of between 28% (considering the maximum exposure value in the scenario, $252 \text{ days}\cdot\text{year}^{-1}$) and 40% (considering the 95th percentile value of the distribution, $212 \text{ days}\cdot\text{year}^{-1}$). The random values generated ranged between 36 and $252 \text{ days}\cdot\text{year}^{-1}$ (Annex C).

Regarding the population variables of body weight (BW) and life expectancy at birth (AT), both start from normal distributions, so they have a higher degree of symmetry and less pronounced tails. The BW variable represents the variability between children aged 0.5 and 6 years, and has centring parameters that closely match the value proposed by the USEPA (2014a, b) of 15 kg. However, the averaging time for carcinogens, represented by the life expectancy at birth, shows an approximate average value of 80 years, higher than the reasonable maximum exposure value of 70 years assumed for the general population in the USEPA (2014a, b) Risk Assessments. The discrepancy in these values may be due to differences in the populations analysed to obtain information or to the increase in life expectancy since the first value was estimated (USEPA 1989). The latest document updating life expectancy at birth proposes the use of a value very close to that obtained by our study of 78 years (USEPA 2011b). Consideration of the default AT parameter (70 years) could lead to an increase in risk of approximately 10% in the scenario under study. The range of the random numbers generated spanned 34.2 kg and 6.36 years for BW and AT, respectively (Annex C).

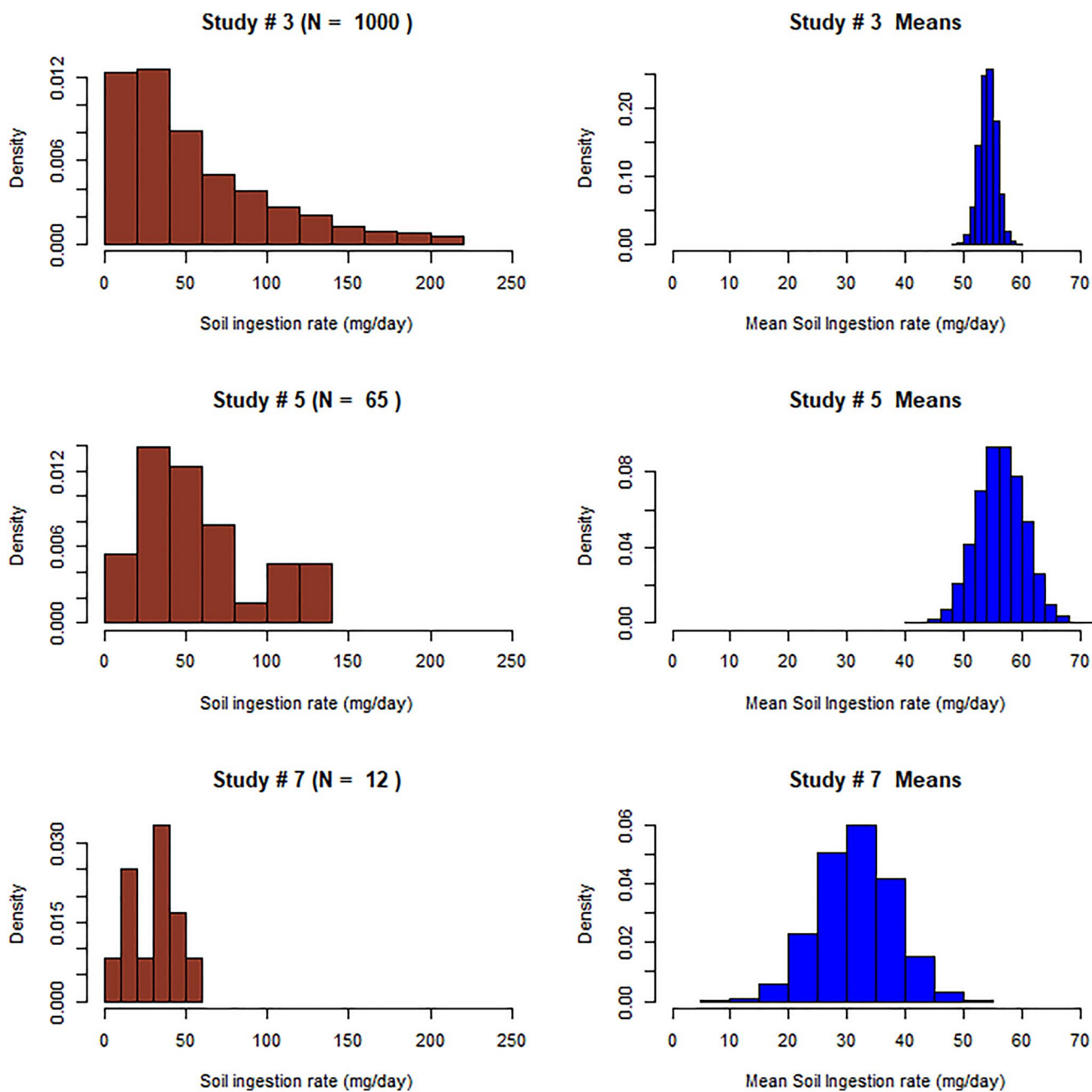


Fig. 2 Examples of mean variability distributions obtained with Monte Carlo simulations. Study numbers correspond to the respective lines in Annex A for the Soil Ingestion Rate variable, respectively:

Study #3 Özkaynak et al. (2011), Study #5 Thompson and Burmaster (1991) and Study #7 (Davis and Mirick (2006)

The As concentration term has been obtained with information from two geochemical investigations in the study area. On the one hand, the work of De Miguel et al. (2002) collected information on the distribution of concentrations of this trace element in the arkose unit of the region of Madrid. On the other hand, the characterisation by Mingot et al. (2011) did not show significant differences between As concentrations in playgrounds and the arkose lithological unit that underlies the city of Madrid.

This lack of differences allows us to assume that the origin of As in playgrounds is geogenic and not anthropogenic, so that both sources of information can be considered together when characterising the variability of the concentration term. USEPA (1992) recommends the use of the upper confidence limit of the mean to estimate the exposure of a receptor at a site. The assumption of the residential scenario implies that contact with soil will occur more frequently in playgrounds close to dwellings, so that

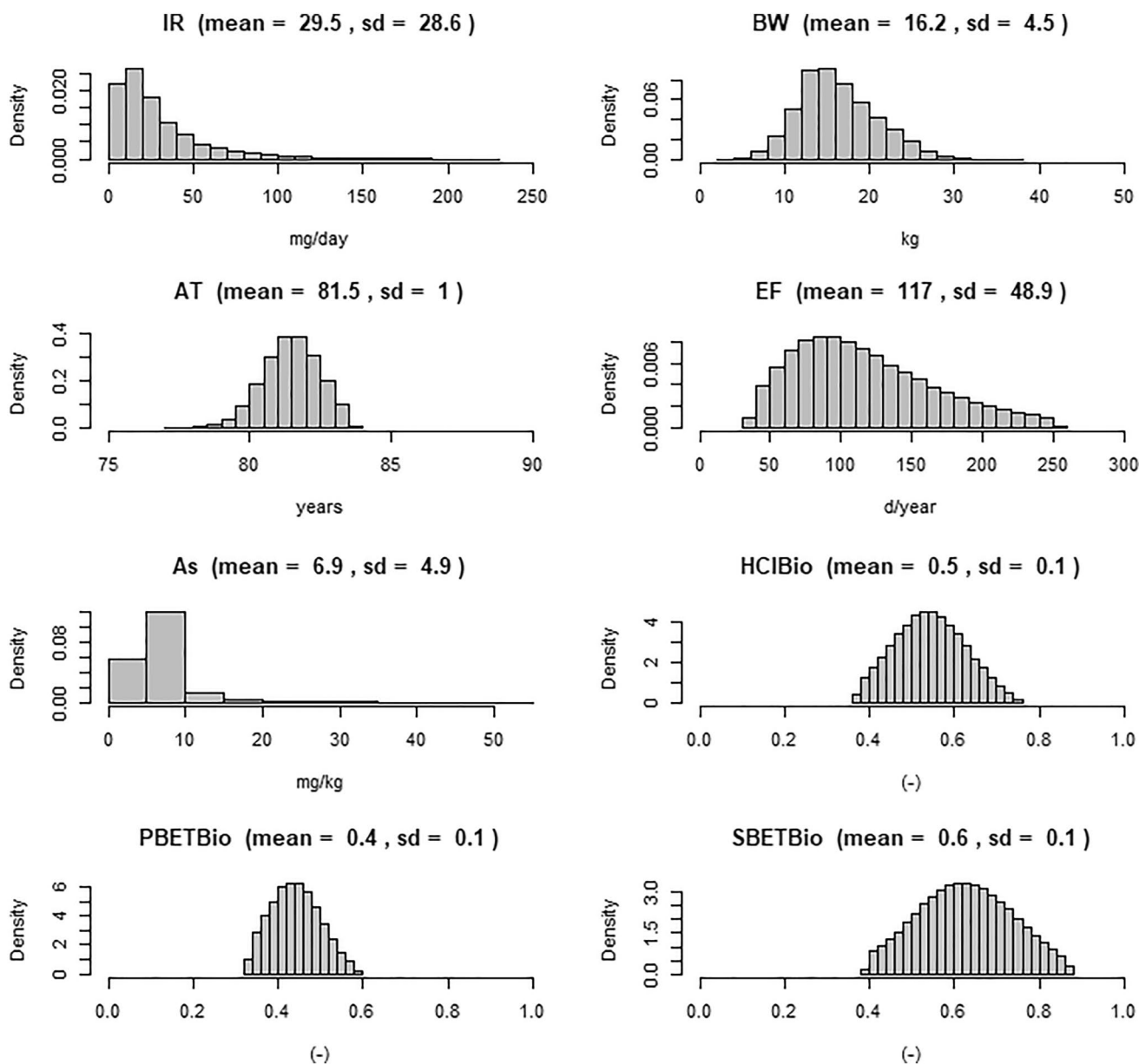


Fig. 3 Histograms of the random variables generated

the variability of the data between playgrounds (indistinguishable from the geological unit of arkose) may be more representative in characterising the variability of exposure in the population than the mean concentrations of all playgrounds in a city. That is, a receptor would be exposed more frequently to the concentrations of a single park. The distribution constructed using the MC-2D method produced a mean value of 6.9 mg kg^{-1} , which is consistent with the means of the two studies considered (6.7 and 6.8 mg kg^{-1} , respectively). However, the 95th percentile of the distribution of As concentrations that results when the high-end values of the natural variability obtained for As concentrations in the region of Madrid are used (Annex C)

is approximately twice as large as the UCL95 calculated by Mingot et al. (2011). This result highlights the importance of performing an unbiased background values characterization while conducting a HHRA in trace element exposure scenarios according to the study by de Lima Brum et al. (2022). It is also interesting to study the impact of considering the bioaccessible fraction of As on the estimate of risk in playgrounds. Based on the results of the distributions generated for the different extraction methods, the concentration term would have an average reduction of 54% with a range between 37 and 73%. Therefore, the increase in the risk estimation produced by considering the natural variability of As contents in the lithological

unit (which ranged between 1.1 and 52.9 mg·kg⁻¹, Annex C), would be counterbalanced by the introduction of As bioaccessibility in the calculations.

It should be noted that some of the random variables presented have been generated through a selection of bibliographic sources that may have a certain level of bias. While the selection of different sources of information to characterise the input variables may provide a good representation of their true heterogeneity, it should be noted that the specific relevance (i.e., site-specificity) of each data source to the exposure scenario of interest has not been quantitatively considered in the model. In other words, a small, site-specific study may have a major relevance in the model output, even if it does not capture the true variability of a given parameter, and therefore, whenever possible, HHRAs should consider local studies to characterise exposure.

Lastly, if we consider the variability reported by different studies in the estimation of the reasonable maximum exposure values (RME) for the different exposure variables and the 95th percentile for the total As concentration, the following values are obtained: IR = 86 mg·day⁻¹, C = 15.06 mg·kg⁻¹; EF = 252 days·year⁻¹, BW = 16 kg, AT = 80 years, ED = 5.5 years. Recalculating the carcinogenic risk with these values and assuming 100% bioaccessibility, a value of 5.76·10⁻⁶ is obtained. This is approximately half of the risk estimated from the default exposure parameters. Therefore, this correction would have a similar impact to that of considering bioaccessibility in the risk estimation. If we use the UCL95 obtained in the study of Mingot et al. (2011) for playgrounds in Madrid (C = 7.27 mg kg⁻¹) instead of the 95th percentile, the risk would be reduced to one fourth of the original value (2.78·10⁻⁶), which would be twice the change resulting from the application of the average bioaccessibility of arsenic.

Probabilistic Risk Assessment

To study the impact of the variability and uncertainty of exposure parameters on risk indeterminacy, the generated random data vectors were introduced into the carcinogenic risk equations to obtain risk distributions for the target population. The result of this process considering total As concentrations in soil and bioaccessible fractions obtained with three analytical methods is shown in Fig. 4 and summarized in Annex D. The results show that there is a possibility of exceeding the acceptability threshold of 10⁻⁵ when considering the total and the SBET- and HCl-bioaccessible concentrations but with very low probabilities of occurrence (0.038, 0.003, and 0.002%, respectively). In all cases, 95% of the population would be at least one order of magnitude below the predictions made with the default exposure parameters in the deterministic risk assessment (Table 1). A comparison of the results between the risk distributions found for total and

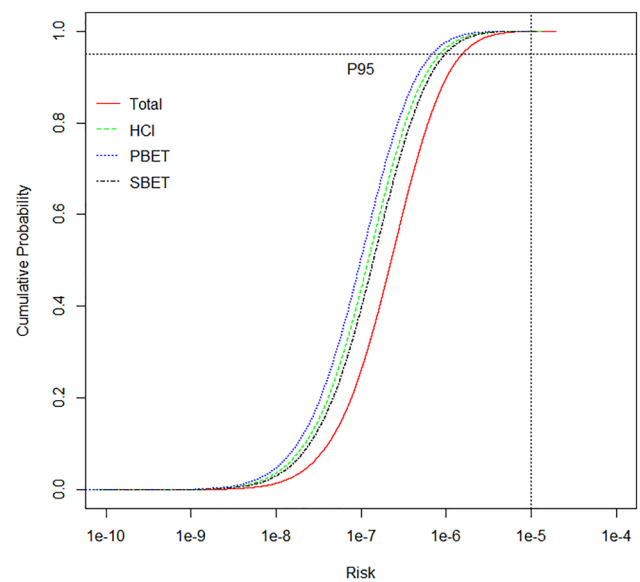


Fig. 4 Cumulative probability distributions obtained for carcinogenic risk considering total As concentrations and bioaccessibility percentages obtained with different methods

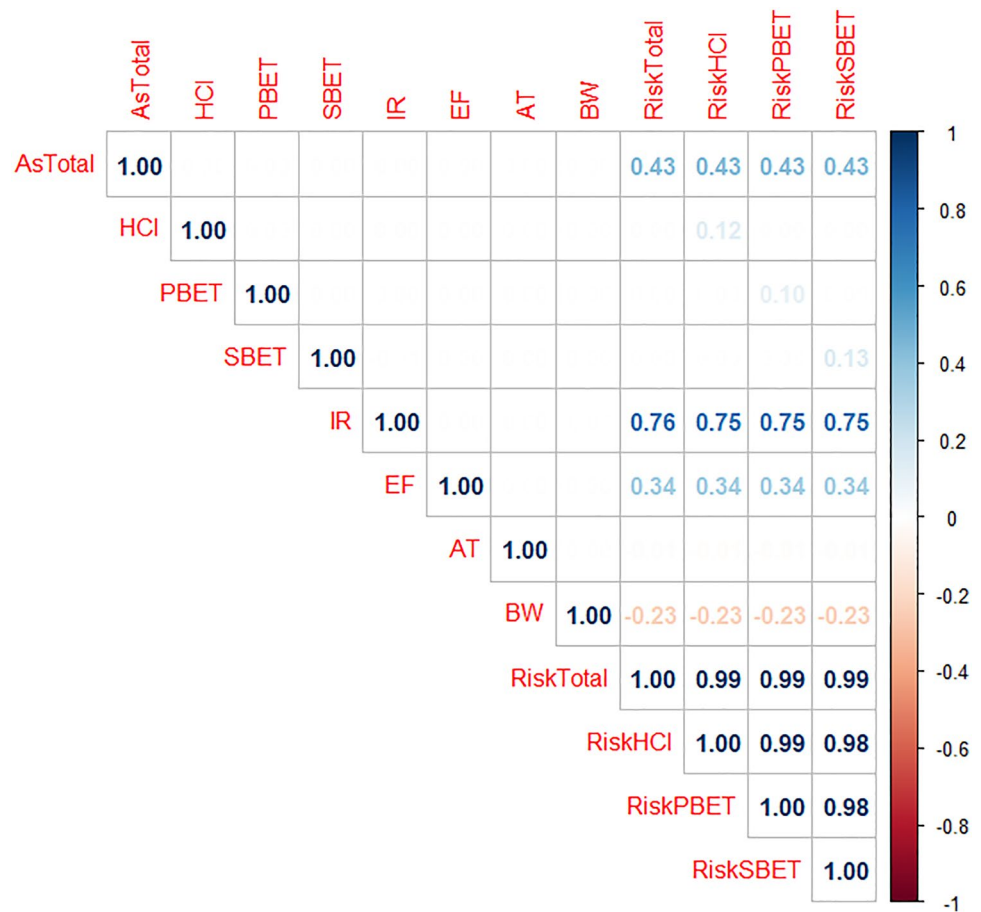
bioaccessible concentrations shows that their ratios are equal to the mean bioaccessibility fractions (0.61, 0.55 and 0.43 for the SBET, HCl and PBET methods, respectively). This result is expected given the symmetry of the indeterminacy of these variables around the mean.

Sensitivity and Significance Analysis

Figure 5 presents in matrix form the Spearman correlation coefficients between the exposure factors produced through the 2D Monte Carlo simulations and the risk distributions obtained using the *cor.sign* function (Filzmoser 2015). This analysis allows identification of the variables that most condition the risk distribution: soil intake rate (IR) and total arsenic concentration (AsTotal) and exposure frequency (EF). Among all variables, the risk produced was most sensitive to the soil intake rate with a Rank Correlation higher than 0.6 (Burmester and Anderson 1994; Harris and Horn 1996). This correlation can be explained by the heavily-tailed distribution generated for the IR parameter, a product of the indeterminacy between studies on this exposure factor (especially at the high percentiles of the distributions). This phenomenon is repeated for the variability of the total arsenic concentration, whose PDF is generated from heavily-tailed log-normal distributions and, to a lesser extent, with the EF parameter. These results agree with the ranges of the generated random variables described in “Random variable generation” and Annex C.

The exposure parameters constructed from normal distributions (body weight (BW) and the different

Fig. 5 Spearman correlation coefficients obtained between the generated random variables and the risk distributions



bioaccessible fractions (HCl, PBET, and SBET)) showed lower correlation coefficients. While in the case of AT the correlation coefficient is null due to the low dispersion of the data (given the great statistical power of the information sources), the rank correlation increases in the case of BW, whose variability is constructed from numerous normal distributions that add complexity to the generated PDF. An intermediate between those of BW and AT is the distributions of bioaccessible fractions which, despite their symmetry, present a sufficient range of variation to generate correlations greater than zero (although these are still weak).

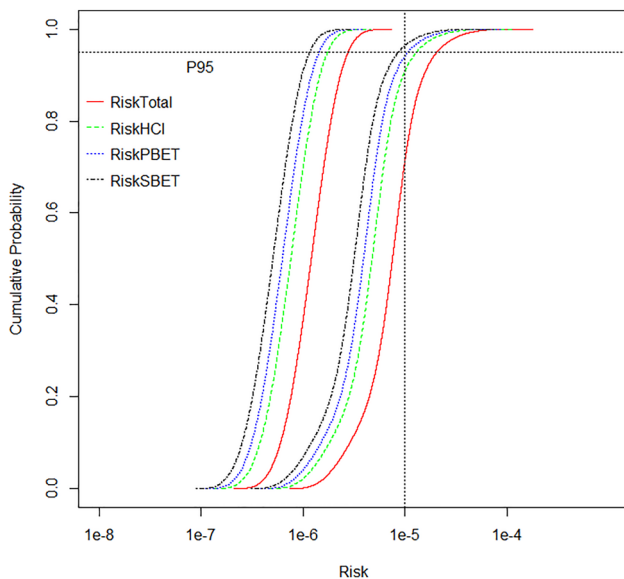
Interestingly, in all simulations, the spread of risk data is more dependent on the geochemical variability of As, IR, EF, and even BW, than on the analytical indeterminacy of the bioaccessibility percentage (irrespective of the method). This result suggests that the analytical indeterminacy of bioaccessible methods does not have a significant impact on the uncertainty of risk compared to the natural variability of As concentrations in the geological unit in which the city of Madrid is located, the uncertainty in the characterisation of the soil accidental intake rate, the variability in the patterns of use of playgrounds, or the variability of body weight within the population.

Based on the results of the sensitivity analysis and considering the most influencing variables (total As concentration, IR, and EF), two groups of simulations are proposed to establish the impact of their variability on risk uncertainty (importance analysis) (Table 2). Simulation Group #1 contains the lower bounds of variation of the exposure variables (representing the lower bounds of risk variability), while Simulation Group #2 consists of the upper bounds of variation of the input variables (generating the upper bounds of risk indeterminacy). Both groups contain four simulations in total, considering (i) the total concentration of As ($BA = 1$) and (ii) the average bioaccessibility percentages obtained by the three digestion methods considered (HCl, PBET and SBET).

In the case of the total concentration, the lower limit for the sensitivity analysis has been set at the UCL95 value obtained in the study by Mingot et al. (2011), while the upper limit has been set as the PDF obtained from the application of the MC-2D method. For the soil ingestion rate variable, the 95th percentile of the PDF generated from the MC-2D study has been set as lower limit, while the upper limit has been set as the deterministic parameter used by default in the BRA (USEPA 2014a, b). Regarding the frequency of exposure, the PDF generated from

Table 2 Groups of simulations proposed for the significance analysis (MC2D PDF: Probability Distribution Function generated from the 2D Monte Carlo simulations; see “Random variable generation”)

Simulation group	As total Concentration (mg·kg ⁻¹)	IR (mg day ⁻¹)	EF (day·year ⁻¹)	BW (kg)	AT (year)
#1	UCL95 = 7.27	86	MC2D PDF	MC2D PDF	MC2D PDF
#2	MC2D PDF	200	350	MC2D PDF	MC2D PDF

**Fig. 6** Results of simulations carried out to analyse the importance of the variability of exposure factors

local information has been considered for group #1, while for group #2, the value of the reasonable maximum exposure for a residential scenario has been included (USEPA 2014a, b). In both groups, the PDFs generated by MC-2D for body weight and averaging time have been included.

The results of the simulations are shown in Fig. 6. As expected, the CDFs of Group #2 are shifted to the right with respect to those of Group #1. In Group #1 simulations the probabilities of exceeding the threshold value of 10^{-5} are 0 in all cases, while for Group #2 it is 28%, 9%, 6% and 4% considering the total and bioaccessible concentrations obtained with the SBET, HCl and PBET methods, respectively. Focussing on the 95th percentiles of the distributions generated for the total As concentrations in both simulations ($2.75 \cdot 10^{-6}$ for Group #1 and $2.05 \cdot 10^{-5}$ for Group #2), we can observe a seven-fold increase in the risk when considering the default exposure parameters of USEPA (2014a, b). The omission of the bioaccessibility parameter implies an overestimation of risk of up to 2.4 times in the two combinations of exposure factors considered.

Conclusions

This work illustrates the usefulness of an algorithm based on 2DMC for the generation of random variables that consider the effect of the variability and uncertainty of the exposure parameters (including different local and bibliographic sources of information) in the probabilistic analysis of risk. The results obtained in this particular case study reflect the high impact on the risk estimates of the uncertainty associated with the accidental soil and dust ingestion rate parameter, and of the natural variability of the background concentration of As in the geological unit studied. The uncertainty in the definition of the scenario, represented in this study by the differences between the exposure frequency estimated for local receptors versus the recommended exposure frequency for a residential scenario, has been shown to have a significant impact on the indeterminacy of risk. Although numerous studies have focussed their efforts on reducing the uncertainty of the concentration term by introducing a bioaccessibility factor, the results obtained in this work show that bioaccessibility has less impact on the indeterminacy of risk than that introduced with the use of bibliographic exposure parameters, the variability of the population's habits, and the variability of the concentration of the toxic substance in soil. Therefore, and with the aim of reducing uncertainties in the HHRA, the results of this study highlight the importance of increasing efforts in the geochemical characterization of the site, the behavioral patterns of the exposed population and, as far as possible, the contact rates with the contaminated environment (without neglecting the direct impact of the determination of bioaccessibility in the reduction of uncertainty in the risk characterization, which could be more pronounced in other sites and for other contaminants).

Although the Probabilistic Risk Analysis methodology has shown its usefulness in characterising the variability of risk within the study population, the widely used deterministic approach still shows advantages in terms of the generalizability of its implementation, making the results obtained for different scenarios easily comparable. This facilitates the work of risk managers and the application of homogeneous criteria in decision making. Although the characterisation of certain variables at the local level may be difficult (e.g., accidental soil and dust ingestion rate), risk analysts should proceed with the updating of reasonable maximum exposure parameters to reduce uncertainties in their results. The methods applied in this work may be useful for this purpose.

Appendix A. Bibliographic information used for the construction of random variables

In this appendix, the list of information sources and input parameters used in the simulations are shown in tabular form

	Dist	n_i	μ_i	σ_i	\min_i	\max_i	Reference
As total concentration (CS) $\text{mg}\cdot\text{kg}^{-1}$	LN	32	6.64	1.24	4.4	11.5	Mingot et al. (2011)
	LN	64	6.68	8.11	1.1	42.9	De Miguel et al. (2002)
SBET-Bioaccessibility (BA) unitless	N	32	0.62	0.13	0.40	0.87	Mingot et al. (2011)
HCl-Bioaccessibility (BA) unitless	N	32	0.53	0.09	0.37	0.75	
PBET-Bioaccessibility (BA) unitless	N	32	0.43	0.07	0.33	0.59	
Ingestion rate (IR) $\text{mg}\cdot\text{day}^{-1}$	LN	200,000	20	26	0	64	(Wilson et al., 2013)
	LN	200,000	23	32	0	75	
	LN	1000	67.61	90.62	4.86	200	(Özkaynak et al., 2011)
	LN	216	25.5	15.5	0	79.4	(Stanek et al. 2012)
	LN	65	91	126	13	143	(Thompson and Burmaster, 1991)
	LN	64	31	31	0	137	(Stanek et al., 2001)
	LN	12	36.7	35.4	0	107.09	(Davis and Mirick, 2006)
	LN	12	38.1	31.4	0	95	
Exposure frequency (EF) $\text{day}\cdot\text{year}^{-1}$	LN	36	136	66.16	36	252	De Miguel et al. (2007)
Body weight (BW) kg	N	2974	3.31	0.43	6.2	10.6	Carrascosa et al. (2008b, 2008a)
	N	233	6.25	0.9	7.3	12	
	N	214	8.1	1.09	8.5	13	
	N	213	9.32	1.13	8.3	14	
	N	169	10.37	1.12	10	15.8	
	N	166	11.22	1.4	9.6	15.5	
	N	149	12.08	1.45	10.2	16	
	N	153	12.59	1.34	11.4	17.5	
	N	182	12.9	1.42	11.5	21.7	
	N	263	14.14	1.6	12.4	24.2	
	N	508	15.4	2	13.2	24.7	
	N	463	16.32	2.61	14.5	28.3	
	N	469	17.5	2.63	15	31.4	
	N	438	18.58	2.83	15.8	31.8	
	N	370	19.71	3.48	16.8	32.8	
	N	363	21.41	3.65	6	9.4	
	N	332	22.37	3.751	6.5	11.8	
	N	2822	3.18	0.41	7.9	11.8	
	N	233	5.64	0.8	8	12.8	
	N	205	7.4	0.92	8.4	13.9	
	N	186	8.71	1.15	9.7	14.9	
	N	175	9.73	1.08	9.8	16.4	
	N	138	10.32	1.14	10.5	17.8	
	N	125	11.12	1.3	11	18.3	
	N	106	11.82	1.31	12	21.4	

	Dist	n_i	μ_i	σ_i	\min_i	\max_i	Reference
	N	114	12.49	1.4	12.3	23	
	N	193	13.61	1.52	13.8	24.5	
	N	474	14.55	1.68	14.5	26.4	
	N	382	15.61	2.28	15.4	31.4	
	N	420	16.8	2.41	16.1	32.2	
	N	358	18.06	2.55	10.73	15.54	
	N	407	19.37	3	12.69	18	
	N	370	20.65	3.75	14.4	21.98	
	N	336	22.15	3.88	17.29	29.4	
	N	411	13.02	1.26	10.05	15.27	Durá Travé (2012)
	N	364	15.34	1.49	11.81	18	
	N	360	17.75	2.06	13.52	21	
	N	365	22.6	3.13	16.94	28.32	
	N	371	12.38	1.41	6.2	10.6	
	N	330	14.71	1.64	7.3	12	
	N	339	16.97	2.1	8.5	13	
	N	359	21.79	2.88	8.3	14	
Averaging time (AT) year	N	46,745,807	81.58	1.01	0	120	Goerlich Gisbert (2012)

Dist probability distribution Function reported by the literature source (*N* Normal Distribution, *LN* Log-normal distribution)

Appendix B. Exposure factors used for the Baseline Risk Assessment of the risk for an urban recreational scenario

In this appendix, the list Exposure factors employed in the Baseline Risk Assessment are shown in tabular form

Exposure factor	Site-specific	Generic	Reference
As UCL95 (CS) $\text{mg}\cdot\text{kg}^{-1}$	7.27		Mingot et al. (2011)
Bioaccessibility (BA) unitless			
SBET	0.63		
HCl	0.52		
PBET	0.42		
Ingestion rate (IR) $\text{mg}\cdot\text{day}^{-1}$		200	USEPA (2014a)
Exposure frequency (EF) $\text{day}\cdot\text{year}^{-1}$		350	
Exposure duration (ED) year		6	
Bodyweight (BW) kg		15	
Average time (for carcinogenic effects) (AT) day		25,550	

Appendix C. Descriptive statistics of the generated exposure variables

This annex contains the descriptive statistics of the distributions of the exposure variables constructed with the 2D Monte Carlo method

Variable	Min	5th percentile	1Q	2Q	Mean	3Q	95 th percentile	Max
IR (mg·day ⁻¹)	0.04	3.3	11.1	20.7	29.5	37.7	86.3	223.8
BW (kg)	3.7	9.50	13.0	15.6	16.2	19.15	24.5	37.9
AT (year)	77.22	79.86	80.83	81.50	81.47	82.15	83.01	83.58
EF (days·year ⁻¹)	36.00	50.44	78.97	108.78	116.82	147.53	211.97	251.99
As (mg·kg ⁻¹)	1.10	1.75	4.65	6.25	6.88	7.71	15.06	52.90
HCIBio (–)	0.37	0.41	0.48	0.54	0.54	0.60	0.68	0.75
PBETBio (–)	0.33	0.35	0.40	0.44	0.44	0.48	0.54	0.58
SBETBio (–)	0.40	0.45	0.55	0.63	0.63	0.71	0.81	0.87

Appendix D. Descriptive statistics of the generated risk distributions

Risk	Min	5th percentile	1Q	2Q	Mean	3Q	95 th percentile	Max
As Total	1.5E-10	2.4E-08	9.5E-08	2.2E-07	4.4E-07	5.0E-07	1.5E-06	2.0E-05
SBET-Bioaccessibility	8.4E-11	1.4E-08	5.9E-08	1.4E-07	2.8E-07	3.1E-07	9.7E-07	1.2E-05
HCl-Bioaccessibility	7.3E-11	1.2E-08	5.1E-08	1.2E-07	2.4E-07	2.7E-07	8.4E-07	1.3E-05
PBET-Bioaccessibility	5.8E-11	1.0E-08	4.1E-08	9.8E-08	1.9E-07	2.2E-07	6.8E-07	8.9E-06

Author Contributions All authors contributed to the study conception and design. The methodology and software were developed by FBP. The formal analysis, investigation and visualization were performed by FBP and HSG. MID performed data curation. The first draft was prepared by FBP and HSG. All authors reviewed and edited the final version of the manuscript. EDM acquired funding and perform the project administration.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature. This study was funded through the CARESOIL–CM [P2018/EMT-4317] research grant of the Regional Government of Madrid (Comunidad de Madrid), and the “ROTATE: CIRCULAR ECOLOGICAL ESSENTIAL & CRITICAL RAW MATERIALS” project [HORIZON-IA 101058651] funded by the Horizon Europe Framework Program (HORIZON) of the European Commission.

Data Availability Detailed code and the database employed in this work are available upon request from Fernando Barrio-Parra (fernando.barrio@upm.es).

Declarations

Competing Interests The authors have no relevant financial or non-financial interests to disclose.

Ethical Approval This study did not involve human or animal subjects. All the data employed was obtained from bibliographic information and previously published research databases, so it did not require ethics approval.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Barrio-Parra F, De Miguel E, Lázaro-Navas S, Gómez A, Izquierdo-Díaz M (2018) Indoor dust metal loadings: a human health risk assessment. *Exposure and Health* 10:41–50. <https://doi.org/10.1007/s12403-017-0244-z>

- Barrio-Parra F, Izquierdo-Díaz M, Dominguez-Castillo A, Medina R, De Miguel E (2019) Human-health probabilistic risk assessment: the role of exposure factors in an urban garden scenario. *Landsc Urban Plan* 185:191–199. <https://doi.org/10.1016/j.landurbplan.2019.02.005>
- Burmester DE, Anderson PD (1994) Principles of good practice for the use of Monte Carlo techniques in human health and ecological risk assessments. *Risk Anal* 14:477–481. <https://doi.org/10.1111/j.1539-6924.1994.tb00265.x>
- Burmester DE, Crouch EAC (1997) Lognormal distributions for body weight as a function of age for males and females in the United States, 1976–1980. *Risk Anal* 4:499–505
- Calabrese EJ, Stanek, EJ Gilbert, CE (1991) Evidence of Soil-Pica Behaviour and Quantification of Soil Ingested. *Hum Exp Toxicol* 10(4):245–249. <https://doi.org/10.1177/096032719101000403>
- Cao S, Duan X, Zhao X, Chen Y, Wang B, Sun C, Zheng B, Wei F (2016) Health risks of children's cumulative and aggregative exposure to metals and metalloids in a typical urban environment in China. *Chemosphere* 147:404–411. <https://doi.org/10.1016/j.chemosphere.2015.12.134>
- Carrascosa A, Fernández García JM, Fernández Ramos C, Ferrández Longás A, López-Siguero JP, Sánchez González E, Sobradillo Ruiz B, Yeste Fernández D (2008a) Estudio transversal español de crecimiento 2008. Parte II: Valores de talla, peso e índice de masa corporal desde el nacimiento a la talla adulta. *An Pediatr* 68:552–569. <https://doi.org/10.1157/13123287>
- Carrascosa A, Fernández JM, Fernández C, Ferrández A, López-Siguero JP, Sánchez E, Sobradillo B, Yeste YD (2008b) Estudios españoles de crecimiento 2008. Nuevos Patrones Antropométricos *Endocrinol y Nutr* 55:484–506. [https://doi.org/10.1016/S1575-0922\(08\)75845-5](https://doi.org/10.1016/S1575-0922(08)75845-5)
- Das S, Jean JS, Kar S (2013) Bioaccessibility and health risk assessment of arsenic in arsenic-enriched soils, Central India. *Ecotoxicol Environ Saf* 92:252–257. <https://doi.org/10.1016/j.ecoenv.2013.02.016>
- Davis S, Waller P, Davis S, Waller P, Buschbom R, Ballou J, White P (1990) Quantitative estimates of soil ingestion in normal children between the ages of 2 and 7 years: Population-based estimates using aluminum, silicon, and titanium as soil tracer elements. *Arch Environ Health* 45:112–122. <https://doi.org/10.1080/00039896.1990.9935935>
- de Lima Brum R, Penteado JO, Ramires PF, Girónes MCR, Mondelongo SP, del Carmen Rubio Armendáriz, M., dos Santos, M., da Silva Júnior, F.M.R., (2022) Recommended guidance and checklist for human health risk assessment of metal(loid)s in soil. *Expo Heal* 14:295–304. <https://doi.org/10.1007/s12403-021-00440-6>
- De Miguel E, Iribarren I, Chacón E, Ordoñez A, Charlesworth SM (2007) Risk-based evaluation of the exposure of children to trace elements in playgrounds in Madrid (Spain). *Chemosphere* 66:505–513. <https://doi.org/10.1016/j.chemosphere.2006.05.065>
- De Miguel E, Mingot J, Chacón E, Charlesworth SM (2012) The relationship between soil geochemistry and the bioaccessibility of trace elements in playground soil. *Environ Geochem Health* 34:677–687. <https://doi.org/10.1007/s10653-012-9486-7>
- De Miguel E, Izquierdo-Díaz M, Gómez A, Mingot J, Barrio-Parra F (2017) Risk assessment from exposure to arsenic, antimony and selenium in urban gardens (Madrid, Spain). *Environ Toxicol Chem* 36:544–550. <https://doi.org/10.1002/etc.3569>
- De Miguel E, Ordoñez A, Barrio-Parra F, Izquierdo-Díaz M, Álvarez R, Mingot J, Charlesworth SM (2019) Bioaccessibility of trace elements in urban environments. *Urban pollution: science and management*. Wiley, Chichester, pp 107–113
- Denys S, Caboche J, Feidt C, Hazebrouck B, Dor F, Dabin C, Floch-Barneaus A, Tack K (2009) Biodisponibilité et bioaccessibilité des métaux et metalloïdes des sols pollués pour la voie orale chez l'homme. *Environnement, Risques & Santé* 8:433–438
- Dong Z, Liu Y, Duan L, Bekele D, Naidu R (2015) Uncertainties in human health risk assessment of environmental contaminants: a review and perspective. *Environ Int* 85:120–132. <https://doi.org/10.1016/j.envint.2015.09.008>
- Dourson ML, Gadagbui B, Griffin S, Garabrant DH, Haws LC, Kirman C, Tohyama C (2013) The importance of problem formulations in risk assessment: a case study involving dioxin-contaminated soil. *Regul Toxicol Pharmacol* 66:208–216. <https://doi.org/10.1016/j.yrtph.2013.03.010>
- Dudka S, Miller WP (1999) Permissible concentrations of arsenic and lead in soils based on risk assessment. *Water Air Soil Pollut* 113:127–132. <https://doi.org/10.1023/A:1005028905396>
- Durá Travé T (2012) ¿Son válidas las curvas y tablas de crecimiento españolas actuales? *Nutr Hosp* 27:244–251. <https://doi.org/10.3305/nh.2012.27.1.5474>
- Filzmoser, P., 2015. StatDA: Statistical Analysis for Environmental Data. R package version 1.6.9. <https://CRAN.R-project.org/package=StatDA>
- Frey HC, Rhodes DS (1996) Characterizing, simulating, and analyzing variability and uncertainty: an illustration of methods using an air toxics emissions example. *Hum Ecol Risk Assess* 2:762–797. <https://doi.org/10.1080/10807039609383650>
- Ge Y, Murray P, Sauvé S, Hendershot W (2002) Low metal bioavailability in a contaminated urban site. *Environ Toxicol Chem* 21:954–961
- Glennec P, Lucas JP, Mandin C, Le Bot B, Le B (2012) French children's exposure to metals via ingestion of indoor dust, outdoor playground dust and soil: contamination data. *Environ Int* 45:129–134. <https://doi.org/10.1016/j.envint.2012.04.010>
- Goerlich Gisbert, F.J., 2012. Esperanza de vida y causas de muerte. Un análisis de descomposición (1975 - 2009), Documentos de Trabajo. Bilbao (SPAIN).
- Guney M, Welfringer B, Repentigny CD, Zagury GJ (2013) Children's exposure to mercury-contaminated soils: exposure assessment and risk characterization. *Arch Environ Contam Toxicol* 65:345–355. <https://doi.org/10.1007/s00244-013-9891-7>
- Hamad SH, Schauer JJ, Shafer MM, Al-Rheem EA, Skaar PS, Heo J, Tejedor-Tejedor I (2014) Risk assessment of total and bioavailable potentially toxic elements (PTEs) in urban soils of Baghdad-Iraq. *Sci Total Environ* 494–495:39–48. <https://doi.org/10.1016/j.scitotenv.2014.06.006>
- Hamby DM (1994) A review of techniques for parameter sensitivity analysis of environmental models. *Environ Monit Assess* 32:135–154. <https://doi.org/10.1007/BF00547132>
- Hammonds JS, Hoffman FO, Bartell SM (1994) An introductory guide to uncertainty analysis in environmental and health risk assessment. US DOE. <https://doi.org/10.2172/10127301>
- Han Q, Wang M, Cao J, Gui C, Liu Yanping, He X, He Y, Liu Yang (2020) Health risk assessment and bioaccessibilities of heavy metals for children in soil and dust from urban parks and schools of Jiaozuo China. *Ecotoxicol Environ Saf*. <https://doi.org/10.1016/j.ecoenv.2019.110157>
- Harris G, Horn R Van (1996) Use of Monte Carlo Methods in Environmental Risk Assessment at the INEL: Applications and issues. Lockheed Idaho Technologies Co., Idaho Falls: 74 p. <https://doi.org/10.2172/266659>
- Iman RL, Helton JC (1988) An investigation of uncertainty and sensitivity analysis techniques for computer models. *Risk Anal* 8:71–90. <https://doi.org/10.1111/j.1539-6924.1988.tb01155.x>
- International Programme on Chemical Safety, 2008. Guidance document on characterizing and communicating uncertainty in exposure assessment, harmonization of approaches to the assessment of risk from exposure to chemicals. Geneva.

- Iribarren I, Chacón E, De Miguel E (2009) A Bayesian approach to probabilistic risk assessment in municipal playgrounds. *Arch Environ Contam Toxicol* 56:165–172. <https://doi.org/10.1007/s00244-008-9161-2>
- Izquierdo-Díaz M, De Miguel E, Ortega MF, Mingot J (2015) Bioaccessibility of metals and human health risk assessment in community urban gardens. *Chemosphere* 135:312–318. <https://doi.org/10.1016/j.chemosphere.2015.04.079>
- Jager T, Vermeire TG, Rikken MG, van der Poel P (2001) Opportunities for a probabilistic risk assessment of chemicals in the European Union. *Chemosphere* 43:257–264. [https://doi.org/10.1016/S0045-6535\(00\)00087-4](https://doi.org/10.1016/S0045-6535(00)00087-4)
- Jang HK, Kim JY, Lee JK (2009) Radiological risk assessment for field radiography based on two dimensional Monte Carlo analysis. *Appl Radiat Isot* 67:1521–1525. <https://doi.org/10.1016/j.apradiso.2009.02.076>
- Jiménez-Oyola S, Chavez E, García-Martínez MJ, Ortega MF, Boloño D, Guzmán-Martínez F, García-Garizabal I, Romero P (2021) Probabilistic multi-pathway human health risk assessment due to heavy metal(loid)s in a traditional gold mining area in Ecuador. *Ecotoxicol Environ Saf*. <https://doi.org/10.1016/j.ecoenv.2021.112629>
- Li D, Li L (2021) Human chemical exposure from background emissions in the United States and the implication for quantifying risks from marginal emission increase. *Toxics*. <https://doi.org/10.3390/toxics9110308>
- Li Q, Li F, Xiao MS, Cai Y, Xiong L, Huang JB, Fu JT (2018) Bioaccessibility and human health risk assessment of lead in soil from Daye City. *IOP Conf Ser*. <https://doi.org/10.1088/1755-1315/108/4/042116>
- Liang CP, Jang CS, Chen JS, Wang SW, Lee JJ, Liu CW (2013) Probabilistic health risk assessment for ingestion of seafood farmed in arsenic contaminated groundwater in Taiwan. *Environ Geochem Health* 35:455–464. <https://doi.org/10.1007/s10653-012-9507-6>
- Ljung K, Oomen A, Duits M, Selinus O, Berglund M (2007) Bioaccessibility of metals in urban playground soils. *J Environ Sci Heal Part A* 42:1241–1250. <https://doi.org/10.1080/10934520701435684>
- Martínez-Sánchez MJ, Martínez-López S, Martínez-Martínez LB, Pérez-Sirvent C (2013) Importance of the oral arsenic bioaccessibility factor for characterising the risk associated with soil ingestion in a mining-influenced zone. *J Environ Manage* 116:10–17. <https://doi.org/10.1016/j.jenvman.2012.11.009>
- De Miguel, E., Callaba, A., Arranz, J.C.J.C., Cala, V., Chacón, E., Gallego, E., Alberruche, E., Alonso, C., Fernández-Canteli, P., Iribarren, I., Palacios, H., Fdez.-Canteli, P., 2002. Determinación de niveles de fondo y niveles de referencia de metales pesados y otros elementos traza en suelos de la Comunidad de Madrid, Medio ambiente. Terrenos contaminados no2. Instituto Geológico y Minero de España.
- Millard SP (2013) *EnvStats An R package for environmental statistics*. Springer, New York
- Mingot J, De Miguel E, Chacón E (2011) Assessment of oral bioaccessibility of arsenic in playground soil in Madrid (Spain): a three-method comparison and implications for risk assessment. *Chemosphere* 84:1386–1391. <https://doi.org/10.1016/j.chemosphere.2011.05.001>
- Moya J, Philips L (2014) A review of soil and dust ingestion studies for children. *J Expo Sci Environ Epidemiol* 24:545–554
- Novomestky F, Nadarajah S (2016) *truncdist: Truncated Random Variables*. R package version 1.0-2. <https://CRAN.R-project.org/package=truncdist>
- Oomen AG, Rompelberg CJM, Bruil MA, Dobbe CJG, Pereboom DPKH, Sips AJAM (2003) Development of an in vitro digestion model for estimating the bioaccessibility of soil contaminants. *Arch Environ Contam Toxicol* 44:281–287. <https://doi.org/10.1007/s00244-002-1278-0>
- Oregon Department of Environmental Quality, ODEQ (1999) *Guidance for Use of Probabilistic Analysis in Human Health Risk*. Oregon Department of Environmental Quality, Waste Management & Cleanup Division, Cleanup Policy & Program Development Section. Portland, Oregon: 158 p. <https://digital.osl.state.or.us/islandora/object/osl:21017>
- Penteado JO, de Lima Brum R, Ramires PF, Garcia EM, dos Santos M, da Silva Júnior FMR (2021) Health risk assessment in urban parks soils contaminated by metals, Rio Grande city (Brazil) case study. *Ecotoxicol Environ Saf*. <https://doi.org/10.1016/j.ecoenv.2020.111737>
- Pouillot R, Delignette-Muller ML (2010) Evaluating variability and uncertainty separately in microbial quantitative risk assessment using two R packages. *Int J Food Microbiol* 142:330–340. <https://doi.org/10.1016/j.ijfoodmicro.2010.07.011>
- R Core Team (2022) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Rajasekhar B, Nambi IM, Govindarajan SK (2018) Human health risk assessment of ground water contaminated with petroleum PAHs using Monte Carlo simulations: a case study of an Indian metropolitan city. *J Environ Manage* 205:183–191. <https://doi.org/10.1016/j.jenvman.2017.09.078>
- Rasmussen PE, Beauchemin S, Nugent M, Dugandzic R, Lanouette M, Chénier M (2008) Influence of matrix composition on the bioaccessibility of copper, zinc, and nickel in urban residential dust and soil. *Hum Ecol Risk Assess an Int J* 14:351–371. <https://doi.org/10.1080/10807030801934960>
- Ruby MV, Schoof R, Brattin W, Goldade M, Post G, Harnois M, MOSby E, Casteel SW, Berti W, Carpenter M, Edwards D, Cragin D, Chappell W (1999) Advances in evaluating the oral bioavailability of inorganics in soil for use in human health risk assessment. *Environ Sci Technol* 33:3697–3705. <https://doi.org/10.1021/es990479z>
- Sadeghfam S, Abdi M, Khatibi R, Nadiri AA (2021) An investigation into uncertainties within human health risk assessment to gain an insight into plans to mitigate impacts of arsenic contamination. *J Clean Prod* 311:127667. <https://doi.org/10.1016/j.jclepro.2021.127667>
- Simon TW (1999) Two-dimensional Monte Carlo simulation and beyond: a comparison of several probabilistic risk assessment methods applied to a superfund site. *Hum Ecol Risk Assess* 5:823–843. <https://doi.org/10.1080/10807039.1999.9657762>
- Stanek EJ, Calabrese EJ (1995) Soil ingestion estimates for use in site evaluations based on the best tracer method. *Hum Ecol Risk Assess an Int J* 1:133–157. <https://doi.org/10.1080/10807039509379998>
- Stanek EJ, Calabrese EJ, Xu B (2012) Meta-analysis of mass-balance studies of soil ingestion in children. *Risk Anal* 32:433–447. <https://doi.org/10.1111/j.1539-6924.2011.01673.x>
- Stanfield Z, Addington CK, Dionisio KL, Lyons D, Tornero-Velez R, Phillips KA, Buckley TJ, Isaacs KK (2021) Mining of consumer product ingredient and purchasing data to identify potential chemical coexposures. *Environ Health Perspect* 129:1–14. <https://doi.org/10.1289/EHP8610>
- USEPA (1989) *Risk assessment: guidance for superfund*. In: *Human Health Evaluation Manual (Part A), Interim Final, vol 1*. Office of Emergency and Remedial Response, United States Environmental Protection Agency, Washington DC. https://www.epa.gov/sites/default/files/2015-09/documents/rags_a.pdf
- USEPA (1991) *Risk Assessment Guidance for Superfund. Hum Heal Eval Man (Part B, Dev Risk-based Prelim Remediat Goals) I:54*
- USEPA (1992) *Supplemental guidance to RAGS: calculating the concentration term*. Interim Bull 9285:7–08I

- USEPA (1997) Guiding Principles for Monte Carlo Analysis, Risk Assessment Forum. United States Environmental Protection Agency, Washington DC. <https://www.epa.gov/sites/default/files/2014-11/documents/montecar.pdf>
- USEPA (2001) Risk assessment guidance for superfund (RAGS) volume III - Part A: process for conducting probabilistic risk assessment, appendix B. Off. Emerg. Remedial Response U.S. Environ Prot Agency III:1–385
- USEPA (2011a) Exposure factors handbook 2011 edition EPA/600/R-090/052F. U.S. Environmental Protection Agency, Washington
- USEPA (2011b) Lifetime (Chapter 18). Exposure factors handbook 2011 edition EPA/600/R-090/052F. U.S. Environmental Protection Agency, Washington
- USEPA (2012) Standard Operating Procedure for an In Vitro Bioaccessibility Assay for Lead in Soil. United States Environmental Protection Agency; Washington, DC. <https://semspub.epa.gov/work/HQ/174533.pdf>
- USEPA (2013) ProUCL 5.0 [WWW Document]. URL <http://www2.epa.gov/land-research/proucl-software>
- USEPA (2014a) OSWER Directive 9200.1-120. Human Health Evaluation Manual, Supplemental Guidance: Update of Standard Default Exposure Factors. United States Environmental Protection Agency; Washington, DC. https://www.epa.gov/sites/default/files/2015-11/documents/oswer_directive_9200.1-120_exposurefactors_corrected2.pdf
- USEPA (2014b) Risk Assessment Forum White Paper: Probabilistic Risk Assessment Methods and Case Studies. Office of the Science Advisor; United States Environmental Protection Agency; Washington, DC. <https://www.epa.gov/sites/default/files/2014-12/documents/raf-pra-white-paper-final.pdf>
- USEPA (2017) Update for Chapter 5 of the Exposure Factors Handbook - Soil and Dust Ingestion. Office of Research and Development. United States Environmental Protection Agency Washington, DC. https://www.epa.gov/sites/default/files/2018-01/documents/efh-chapter05_2017.pdf
- USEPA (2022) Integrated risk information system (IRIS) [WWW Document]. URL <http://www.epa.gov/iris/>
- Vadali M, Ramachandran G, Mulhausen J (2009) Exposure modeling in occupational hygiene decision making. *J Occup Environ Hyg* 6:353–362. <https://doi.org/10.1080/15459620902855161>
- Wang Y, Zhu G, Engel B, Wu Y (2020) Probabilistic human health risk assessment of arsenic under uncertainty in drinking water sources in Jiangsu Province. *China Environ Geochem Health* 42:2023–2037. <https://doi.org/10.1007/s10653-019-00476-3>
- Wragg J, Cave M, Taylor H, Basta N, Brandon E, Casteel S, Gron C, Oomen A, Van de Wiele T (2009) Inter-laboratory trial of a unified bioaccessibility testing procedure. British Geological Survey, Nottingham, p 90
- Wu Y, Hoffman FO, Apostoaei AI, Kwon D, Thomas BA, Glass R, Zablotska LB (2019) Methods to account for uncertainties in exposure assessment in studies of environmental exposures. *Environ Heal A Glob Access Sci Source* 18:1–15. <https://doi.org/10.1186/s12940-019-0468-4>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.