WILEY EFSA JOURNAL

Guidelines for Uncertainty Analysis: Application of the respective Documents of EFSA and BfR for Exposure Assessments

Journal:	*EFSA Journal*
Manuscript ID	EFSA-2018-0179
Wiley - Manuscript type:	Supporting publication
Date Submitted by the Author:	26-Jul-2018
Complete List of Authors:	Author, EFSA; European Food Safety Authority, COMMS
EFSA Question Number:	EFSA-Q-2018-00607
Keywords:	uncertainty, uncertainty analysis, scientific assessment
Abstract:	

EXTERNAL SCIENTIFIC REPORT

APPROVED: 27 July 2018

doi:10.2903/sp.efsa.2018.EN-1472

Guidelines for Uncertainty Analysis: Application of the respective Documents of EFSA and BfR for Exposure Assessments

Framework Partnership Agreement GP/EFSA/AMU/2016/01(agreement 2)

Thomas Schendel, Christian Jung, Oliver Lindtner, and Matthias Greiner

German Federal Institute for Risk Assessment

Abstract

Uncertainty analysis is an integral part of risk assessment, therefore institutions like the European Food Safety Authority (EFSA) or the German Federal Institute for Risk Assessment (BfR) have developed uncertainty guidelines. An in-depth comparison of both documents yielded that both documents share the same overall philosophy. They differ mainly in their preferences regarding the employed method for the assessment of uncertainties. While BfR's guideline describes and recommends the use of a qualitative method, EFSA's guidance strongly recommends the quantification of the overall uncertainty. After the comparison of the uncertainty guidelines, they were applied to two case studies. As for the first case study, an exposure assessment regarding the marking of eel with alizarin red S (ARS) and strontium chloride (SC) was selected. Aspects of both guidance documents were used: For the BfR guideline, the default qualitative approach was employed, while for EFSA guidance document, a quantitative approach was used. It should be noted that the step of the overall quantification using an expert knowledge elicitation (EKE) recommended by EFSA was not carried out here for organisational reasons. Since the original risk assessment had very little available time (1-2 weeks), methods were selected which were not time-consuming. A quantitative description of the uncertainty could be obtained. The second case study is an exposure assessment regarding aluminium in cocoa and chocolate. The BfR guideline was mainly employed to identify sources of uncertainties, while the EFSA guidance document, building on the results of the BfR uncertainty analysis, was applied to handle the quantification of these uncertainties. An expert knowledge elicitation (EKE) was employed not only to derive the distribution of one parameter, but also to characterise the overall uncertainty. As a result, the uncertainty of the selected exposure assessment was obtained. The report concludes with an overall evaluation of the two uncertainty guidelines and recommendations for further development are given.

Key words: uncertainty, uncertainty analysis, scientific assessment

Question number: EFSA-Q-2018-00607

Correspondence: thomas.schendel@bfr.bund.de

Acknowledgements: The authors would like to thank following persons: Katrin Blume, Carolin Fechner, Natalie von Götz, Andy Hart, Anna Elena Kolbaum, Laura Martino, Caroline Merten, Olaf Mosbach-Schulz, and Christine Sommerfeld for the support provided to this scientific output.

Suggested citation: EFSA (European Food Safety Authority), Schendel T, Jung C, Lindtner O and Greiner M, 2018. Guidelines for Uncertainty Analysis: Application of the respective documents of EFSA and BfR for exposure assessments. EFSA supporting publication 2018: EN-1472. 113 pp. doi:10.2903/sp.efsa.2018.EN-1472

ISSN: 2397-8325

Summary

The importance of uncertainty analysis for risk assessment has been established over the past five years in food risk assessment. It increases the transparency of the underlying assessment as well as it provides crucial information for decision makers. In order to support risk assessors in their assessment of uncertainties, organizations like the World Health Organization (WHO), ANSES (French Agency for Food, Environmental and Occupational Health & Safety), EFSA, and the BfR have published recommendations regarding uncertainty analysis. The respective guideline of BfR and the guidance document of EFSA are compared in this document.

Both guidelines aim to constitute a complete framework for uncertainty analysis. While the BfR guideline includes risk communication (in contrast to EFSA), the EFSA document additionally provides a toolbox of various methods for uncertainty analysis. A major difference is that the EFSA considers its guidance document to have a broad scope, applicable to all assessments conducted within EFSA, while the BfR guideline limits itself to exposure assessments only.

The purpose of uncertainty analysis is specified similarly: to establish a transparent risk assessment and to empower decision makers to reach an appropriate decision. In this context, the term "uncertainty" is defined for both documents as a lack of knowledge regarding the factors and processes relevant for the particular assessment. Moreover, the term "variability" is introduced and refers to real differences between the members of a population.

The timing of the uncertainty analysis is nearly identical: both guidelines emphasise that uncertainty analysis can indicate whether a refinement of the risk assessment is necessary. Therefore, it should be carried out in parallel with the original assessment. Both guidelines agree on the overall structure of the uncertainty analysis. The EFSA document tends to describe the single steps in more detail and add some additional steps, caused by the more detailed recommendation of methods explained later on. Regarding the identification of uncertainties both guidelines propose question lists. The EFSA's list is (according to their scope) broader, but they also advise to use the question list compiled by the BfR if it seems more appropriate for the assessment at hand.

Both documents distinguish between qualitative (their results are given as verbal expressions or on an ordinal scale) or quantitative (results expressed on a numerical scale) methods for uncertainty analysis. EFSA's guidance strongly recommends that the overall uncertainty is expressed in quantitative terms, since qualitative expressions are ambiguous, they are easily misunderstood, and there is a lack of clear rules of how different types of (qualitative) results are combined. In contrast, the BfR guideline does not demand an overall quantification. For assessing the various sources of uncertainty, it recommends first to use a qualitative approach (especially due to the lower workload), namely a tabular form of assessing uncertainties. If deemed necessary, other (specifically quantitative) methods can be used later on. While the BfR guideline recognizes that the results of quantitative approaches are more precise, due to the increased workload it is only recommended if really needed. Aside from the mentioned qualitative method and some explanation about sensitivity analysis, the BfR guideline does not further introduce other methods. In contrast, the EFSA document lists many qualitative and quantitative methods including explanations as well as discussion of advantages and disadvantages. Moreover, the EFSA guidance describes methods for combining the assessment of uncertainties that have been quantified by calculation (e.g. Monte Carlo simulations). Subsequently, it is described how additional sources of uncertainties, which have not been combined by calculation, can be included in an overall quantification of uncertainty by expert judgement. These topics are not elaborated in the BfR document.

Regarding the refinement of the uncertainty analysis, both guidelines agree that the uncertainty analysis should scale with the needs of the assessment at hand. It suffices using a comparatively low level of refinement, if the resulting uncertainty is deemed acceptable regarding the initial assessment question. However, the definition of a refinement is different: EFSA considers a refinement if formerly

collectively assessed/quantified uncertainties are assessed/quantified individually or if more sophisticated methods are employed. In contrast, the BfR guideline views a refinement as using a higher tier of uncertainty analysis (e.g. if a quantitative approach is used after a qualitative analysis has been performed). Finally, both institutions stress the importance of transparency for documentation.

For a general conclusion, any kind of comparison between the guidelines of the BfR and EFSA is only valid for exposure assessment, since any other field is beyond the scope of BfR's guideline. Given this restriction, both guidelines establish a framework of uncertainty analysis, following basically a comparable philosophy. The main difference is that the EFSA document strongly recommends an overall quantification, while for the BfR guideline qualitative expressions as final results suffice. The BfR uncertainty guideline is limited to exposure assessments, and the case studies selected for this report reflect that.

The first case study dealt with the uncertainty of an exposure assessment regarding the question whether the marking of eel with alizarin red S (ARS) and strontium chloride (SC) poses a risk to consumers. The European eel (*Anguilla anguilla*) is a critically endangered species. One popular measure to increase eel population in inland waters is restocking, where young eels are caught in estuaries and released in inland waters. In order to monitor the efficiency of this measure, these eels can be marked, e.g. with ARS or SC. These marker substances get incorporated in the bones and can later on be detected via microscopy.

The initial risk assessment was characterized by little available time (1-2 weeks) and poor availability of data: the toxicity of the substances in question could not be rated properly and suitable data for conducting an exposure assessment were missing. As a consequence, no exposure estimate was originally determined.

Regarding the uncertainty analysis employing the BfR guideline, the default qualitative approach was used. As a result, the knowledge and data gaps were identified that prevented the exposure estimation of the previous risk assessment. Moreover, the uncertainties of the existing data sources were listed and qualitatively assessed. In conclusion, advice can be given which additional data should be collected. Finally, the results of this qualitative uncertainty analysis constitute a good foundation for a quantitative uncertainty analysis.

The EFSA uncertainty guidance document recommends the quantification of uncertainty. For this case study, it was chosen to build an exposure model with a probabilistic description of the uncertainties. Due to the lack of data, the analysis rested on many assumptions. Nevertheless, important uncertainties could be integrated into the model and assessed quantitatively. If a toxicological threshold value for ARS and SC could be established, the results of the uncertainty analysis would allow assessing the risk of ARS and SC consumption.

The second case study dealt with the uncertainty of an exposure assessment regarding aluminium in chocolate and cocoa. Compared to other foods, these products contain high aluminium concentrations, which motivated the original exposure assessment. The respective report concluded that especially for young children, a significant part of the tolerable weekly intake is exhausted by consumption of cocoa and chocolate.

The uncertainty analysis performed showcased how well the BfR and EFSA guidance complement each other. The BfR uncertainty guideline was mainly used to identify all occurring uncertainties, while the EFSA guidance handled its quantification. In detail, the identified uncertainties were prioritised using a simple sensitivity analysis. As a result it was decided which uncertainties could be handled by a model extension, which parameter needs to be assessed individually, and which need to be determined by an expert knowledge elicitation (EKE). The outcomes of these parts were combined and finally, the remaining uncertainties including newly arising model uncertainties were quantified using an overall

EKE. The results reinforce the findings of the initial BfR exposure assessment and clearly express the magnitude of the associated uncertainties.

The results of the uncertainty analyses of the two case studies have shown that the strengths of each, the BfR and EFSA guidance document, complement each other. The BfR uncertainty guideline has its advantages in identifying sources of uncertainty through the complete path of exposure assessment by using detailed question lists. This kind of detail is (due to the much larger scope) not reached in the EFSA guidance. On the other side, the BfR guideline does not elaborate on how to conduct a quantitative uncertainty assessment. In contrast, the EFSA guidance introduces a large variety of quantitative methods and a complete framework on how to conduct a quantitative uncertainty analysis. The comparative application of the guidelines of EFSA and BfR for the two case studies allowed a direct comparison between a more quantitative and qualitative approach. The results show that a quantitative approach requires time and resources. But it was also found to be difficult to differentiate the degree of impact in the qualitative assessment of the uncertainties.

Table of contents

1. Introduction

The topic of uncertainty analysis in risk assessment has been gaining momentum over the past five years in food risk assessment. There is a general agreement amongst scientific experts that the uncertainties affecting the assessment of risks should be identified, assessed, documented, and communicated. Performing an uncertainty analysis can be a challenging task. Therefore, several organisations have developed guidelines or recommendations to support risk assessors in performing an uncertainty assessment. Examples include the WHO document on uncertainty analysis in exposure assessment (IPCS & IOMC, 2008), a guideline developed by the European Chemicals Agency (ECHA) (ECHA, 2012) and a document prepared by the French Agency for Food, Environmental and Occupational Health & Safety (ANSES, 2016).

Similarly, the European Food Safety Authority (EFSA) and the German Federal Institute for Risk Assessment (BfR) have developed such guidelines. It is important to note, that EFSA calls its respective document guidance; however, if both documents (EFSA and BfR) are mentioned, it will be loosely referred to "guidelines". In order to support their implementation, both institutes started this project with the main goal of testing the practical application of their respective recommendations and develop suggestions towards harmonisation. In a first step, both guidelines are compared in order to identify any possible discrepancies between the two documents. In the next steps, uncertainty analysis is carried out for two case studies applying both guidelines in order to gain practical experience with them. The first case study studies an exposure assessment in the context of the question whether the marking of eel with alizarin red S (ARS) and strontium chloride (SC) may pose a risk to consumers. The uncertainty analysis employing the BfR guideline was a qualitative one, while a quantitative uncertainty analysis was performed using parts of EFSA's guidance document. For the second case study, an exposure assessment of aluminium in cocoa and chocolate was selected. Here both guidelines were employed to support each other – the BfR guideline to identify sources of uncertainties and the EFSA guidance to quantitatively assess these uncertainties. In this framework, expert knowledge elicitation (EKE) was also employed. Finally the experiences of the application of both uncertainty guidelines are summarized in the conclusion.

1.1. Background and Terms of Reference as provided by the requestor

The tasks carried out in this document are part of a larger collaboration project between EFSA and BfR. The work performed for this document is described in the Specific Agreement 2:

"To identify the differences between the EFSA and BfR guidelines on UA, and define working procedures for the case studies as far as such differences could affect the practical implementation or results of UA. "

"To develop at least two application case studies in the BfR (covering qualitative and quantitative approaches, chemical and biological risk assessment) for exemplifying the design, conduct, technical reporting and communication to different target audiences."

This document here is the final report.

This grant was awarded by EFSA to:

Beneficiary: German Federal Institute for Risk Assessment

Grant title: Framework Partnership Agreement, Specific Agreement 2

Grant number: GP/EFSA/AMU/2016/01

2. Methodology and used sources

For reference of the uncertainty guidelines, this document uses the English version of the BfR's guideline document (Heinemeyer *et al.*, 2015). For EFSA's guidance, the respective document was still in the trial phase during this project. Therefore, different parts of this report rely for reference on different versions of EFSA's guidance. For the general comparison of both guidelines, the following draft was used: (EFSA, 2016). For the first case study, it was supplemented by (EFSA, 2017) and for the second case study, the finalised guidance document was employed (Benford *et al.*, 2018).

For comparing the uncertainty guidelines, a list of topic was created. This was deemed necessary as the chapters of both guidelines do not match each other very well. Moreover, some topics are discussed in more than just one place in the documents. The list of topics was generated by going through each document section by section and takes note of all topics in that section.

Secondly, the identified topics were analysed with each other in the following way:

1. The stand-point of the EfSA as well as of the BfR guideline was identified.
2. Both stand-points were compared by listing similarities and differences.
3. If identified, the reasoning for the differences was analysed.

All topics discussed in both guidelines were compared with each other in this way. Additionally, a couple of important topics only mentioned in one guideline were listed. Mostly this concerns the treatment of individually assessed uncertainties in EFSA's document. For these topics, step 3 only consists of analysing the reasoning for not being present in the respective guideline document. All other topics (e.g. introducing the basics of exposure assessment in BfR's guideline) listed only in one document were not compared directly. Finally, at the end of this document, the main differences were summarized, their impact assessed, and consequences for the uncertainty analysis of the two case studies stated.

Regarding the first case study, marking of eel with alizarin red S (ARS) and strontium chloride (SC), an initial risk assessment has been carried out in March 2017 (BfR, 2017a). For this initial risk assessment only a time period of 1-2 weeks was available. Therefore, while not exactly mimicking this time frame, approaches were chosen that could in principle also work in assessments with such a limited timeframe. It should be noted that while the initial risk assessment listed the available knowledge, no exposure model was used and subsequently no exposure was estimated. The reason was the lack of sufficient data. This has profound impact on the uncertainty analysis performed by both guidance documents. The consequences for the uncertainty analysis are stated in the respective guidance's sections.

The second case study, aluminium in cocoa and chocolate, an exposure assessment was carried out in May 2017 (BfR, 2017b). Compared to the first case study, more data and time was available. The use of the two guidelines was such, that they were supplementing each other. The BfR guideline was employed to identify the sources of uncertainty, while the EFSA guidance was used to prioritize these uncertainties, assess them quantitatively with the aim of a quantitative characterisation of the overall uncertainty.

Finally, in the conclusion the experiences of applying the uncertainty guidelines to the two case studies was summarised and recommendations for their use formulated.

3. Comparison of the uncertainty guidelines by EFSA and BfR

3.1. Comparison for each topic

3.1.1. Preliminary Remarks

There are some general differences between the two documents which should be mentioned before a direct comparison of individual topics:

1. EFSA's guidance is much longer (89 pages without annex and extended summary) than the one from BfR (44 pages).
2. While there are some more sections in EFSAs guidance, generally most topics are discussed in more detail in it.
3. EFSA's document provides many more definitions (e.g. 'probability', 'expert judgement') whereas BfR's document seems more to rely on existing definitions (e.g. by WHO).

3.1.2. Purpose of the Document

The purpose refers to the underlying aim of the document. It therefore shapes the overall character of the respective guideline.

3.1.2.1. **BfR Guideline**

The BfR guideline states that the "aim is to ensure uniform methods for the recording, description and assessment of uncertainties in connection with health-related opinions" (p.11) and "[the] guidance document is designed to make it easier to identify the potential and limits of a harmonised methodology for uncertainty analysis" (p.11). Also it shall "provide a template for transparent communication of uncertainties" (p.11). Furthermore, the guideline is set in the context of BfR's general Guidance Document for health risk assessments. The goals of that document – the improvement of the "comprehensibility and coherence of scientific opinions" (p.11), the "support [of] the use of harmonised terminology in the field of risk assessment" and to "ensure the provision of the best-possible advice in the various areas of activity of the BfR" (p.11) are also important for the uncertainty analysis guidance document.

Its main goals are thus the provision of a framework for use in (health) exposure assessments as well as increasing the transparency of the processes within scientific assessments.

3.1.2.2. **EFSA Guidance**

The guidance document aims to "establish a general framework for addressing uncertainty in EFSA" (p.19) and provides "guidance on general principles and a menu (toolbox) of different approaches and methods which can be used to help assessors to systematically identify, characterise, explain and account for sources of uncertainty at different stages of the assessment process" (p.21).

3.1.2.3. **Summary**

The main purpose of both documents is very similar. Both aim to provide a framework in which uncertainty analysis can be performed. The BfR guideline also includes the aim of providing information about transparent communication of uncertainties which is not considered by EFSAs guidance (EFSA Internal Workshop on Guidance on Uncertainty in Scientific Assessments, Parma, 22-

23.06.2017)[1]. EFSA's guidance also intends to supply a toolbox of various methods for addressing uncertainty unlike the BfR document.

3.1.3. Scope of Uncertainty Analysis

'Scope' refers to the area of assessments in which a guidance document considers itself applicable.

3.1.3.1. **BfR Guideline**

The BfR guideline "focusses on the area of exposure assessment" (p.11), acknowledging that other fields are also relevant for uncertainty analysis. It states that it follows the WHO-IPCS guideline on exposure assessment and that there was no guideline on hazard assessment by WHO-IPCS at the time of writing ("Work is currently ongoing at the level of WHO-IPCS on the development of a similar guidance document for uncertainty analysis in hazard assessments", p.11). It considers a revision as soon as the WHO work is finished. In the current form the BfR guideline is thus only applicable to uncertainty analysis in the context of exposure assessments.

3.1.3.2. **EFSA Guidance**

EFSA's guidance document is a "harmonised, but flexible framework that is applicable to all areas of EFSA, all types of scientific assessment, including risk assessment, and all types of uncertainty affecting scientific assessment" (p.20). This scope includes exposure assessment, hazard assessment as well as other types of assessments within the mandate of EFSA. The guidance document stresses that "Uncertainties in decision-making, and specifically in risk management, are outside the scope of EFSA and of this Guidance, as are uncertainties in the framing of the question for scientific assessment" (p.21).

3.1.3.3. **Summary**

The scopes of the documents do not match. The BfR guideline is applicable for exposure assessment while the EFSA guidance includes all of its assessments. As EFSA's scientific assessments can include exposure assessments, the scope of BfR guideline is included in the one of the EFSA.

3.1.4. Definition of Uncertainty; Distinction between Variability and Uncertainty

A very important factor for a guideline concerned with uncertainty analysis is the definition of the term 'Uncertainty'. An often used concept in uncertainty analysis is also the distinction between the terms of 'Variability' and 'Uncertainty'.

3.1.4.1. **BfR guideline**

The BfR guideline references heavily the WHO-IPCS document on uncertainty and data quality in exposure assessment (IPCS & IOMC, 2008; WHO-IPCS, 2004), e.g. by following its guiding principles of uncertainty analysis from that document (cp. p. 12). In the WHO-IPCS document uncertainty is defined as "imperfect knowledge concerning the present or future state of an organism, system, or (sub)population under consideration" (p.103) and as "lack of knowledge regarding the 'true' value of a quantity, lack of knowledge regarding which of several alternative model representations best describes a system of interest or lack of knowledge regarding which probability distribution function and its specification should represent a quantity of interest" (p.103).

[1] But will be addressed in a separate document.

That use of uncertainty by the BfR guideline is further supported by a statement in the foreword, where a situation is described in which "[the] scientists at the BfR are confronted quite often with a situation that they either do not have access to all relevant data, in which the relevant variables have not been analysed, or in which the necessary information is not adequately documented in the available literature" (p. 7).

In respect of distinguishing between variability and uncertainty, the BfR guideline also introduces the terms of 'indeterminacy' and 'difference':

- **Indeterminacy** is used as a general term for the three different ways it can occur, by difference, variability and uncertainty (cp. p.14).
- **Uncertainty** is defined as "lack of knowledge regarding all the factors that influence exposure or the health risks" and "that part of the indeterminacy that occurs when specifying a fixed variable in absence of sufficient knowledge" (p.14).
- **Variability** is "that part of the indeterminacy in the specification of a variable that results from the fact that a variable is observed under different conditions. This generally refers to existing differences between individuals, and variation in time and space. Variability describes a property of the population. It has to be described and it cannot be reduced based on knowledge" (p.14).
- **Difference** occurs when a variable is taken into account differently in "separate models or model stratifications" (p.14). This occurs if due to an examination of the overall variability, different subpopulations are defined and especially different models for each subpopulation are generated. The different models (as well as their different parametrisation) would be referred to difference in terms of the BfR guideline.

3.1.4.2. **EFSA Guidance**

The EFSA guidance gives an explicit statement of the definition of uncertainty:

"In this document, uncertainty is used as a general term referring to all types of limitations in available knowledge that affect the range and probability of possible answers to an assessment question. Available knowledge refers here to the knowledge (evidence, data, etc.) available to assessors at the time the assessment is conducted and within the time and resources agreed for the assessment" (p. 20).

It draws this definition by analysing other documents or groups where uncertainty has been defined or used, including the definition by WHO-IPCS.

Regarding the distinction of variability and uncertainty, the terms are defined in following way:

- **Variability** "is a property of the real world, referring to real differences between the members of a population of real-world entities" (p. 31). Population is to be interpreted in a very broad way, as far as to basically any group of entities. Variability cannot be altered by further research (cp. p. 31).
- **Uncertainty** in this context is used in the exact same definition as given above.

The guidance document also emphasises that these two concepts are closely related, e.g. that a lack of knowledge about a variability is an uncertainty which needs to be taken into account.

3.1.4.3. **Summary**

Although the definition of uncertainty in the BfR guideline is given less prominence than in the EFSA guidance, both documents include a definition of the term. They both agree that uncertainty is caused

by a lack of knowledge regarding any factor influencing the particular risk assessment. This is not surprising since the BfR and the EFSA document are influenced by similar sources (e.g. the WHO-IPCS documents).

Subsequently, both guidelines narrow their scope to questions regarding the assessment at hand, be it all types of assessments within EFSAs mandate or exposure assessment[2] in the case of the BfR guideline. The latter definition of uncertainty is a bit narrower, but this directly follows out of the narrower scope of the BfR guideline. However, it could easily be extended to other types of assessment if needed without changing its core meaning (regarding the lack of knowledge and the assessment at hand).

With reference to variability, both guidelines use very similar definitions. Both agree that it relates to differences of a population occurring in the real world which cannot be reduced by further knowledge.

The BfR guideline additionally defines the term difference.

3.1.5. Purpose of Uncertainty Analysis

In the following, the purpose of both guidelines is compared.

3.1.5.1. **BfR Guideline**

The BfR guideline states in its foreword that "the documentation of existing uncertainties with regard to the status of knowledge and the achievable detail with which these questions can be handled is therefore a matter of good scientific practice and transparency"(p.7). Further down it adds on the aims of uncertainty analysis:

"The aim of an uncertainty analysis is to ensure increased transparency regarding all elements of risk assessment and exposure assessment. In particular, uncertainty analysis should enable decision-makers, stakeholders (interested parties) and the public to gain a better understanding of risk assessment´s content. It should empower them to make their own evidence -based decisions" (p.9).

Also, it considers a "health-related opinion […] incomplete that does not include an adequate description and analysis of uncertainties" (p.11).

Uncertainty analysis therefore mainly serves as a tool for adding transparency to the process of risk assessment. It should also enable decision makers to come to a more informed decision on the management of risks.

3.1.5.2. **EFSA Guidance**

The EFSA guidance document also names transparency as a motivation for performing uncertainty analysis:

"For reasons of transparency […], the assessments must say what sources of uncertainty have been identified and what their impact on the assessment outcome is. This must be reported clearly and unambiguously" (p.21).

It also mentions that the contents and results of an uncertainty analysis are relevant for the decision maker: "Thus, in general, assessors are responsible for characterising uncertainty and decision-makers are responsible for resolving the impact of uncertainty on decisions" (p.23).

[2] The BfR definition also includes lack of knowledge about health risks.

3.1.5.3. **Summary**

The question regarding the purpose of uncertainty analysis is answered in the same way by both guidance documents: It shall increase the transparency of an assessment and furthermore allow decision makers to be able to resolve the impacts of uncertainty on risk management.

3.1.6. Roles of the Risk Assessor and Decision Maker

Both guidelines are intended to be used by risk assessors. Therefore, it is important to compare how both guidelines separate the roles of risk assessor and decision maker.

3.1.6.1. **BfR Guideline**

The BfR guideline does not explicitly state the tasks of the risk assessor and decision makers as such, but gives some indication about their roles in some parts of the document. One example was already mentioned in section 3.4.1, stating that uncertainty analysis should allow decision makers to make fact-based decisions. Another example which emphasises the distinction between risk assessment (including uncertainty assessment) and risk management can be found on page 10:

"The details of findings of an uncertainty analysis, as proposed in this guidance document, must be further summarised for communication with risk managers […]".

It can be concluded, that the BfRs guideline considers the identification and assessment of uncertainties as well of their documentation and communication as job of the risk assessor. This is further underlined by providing information to the assessor on how to address these tasks. Thus, the resolution of the impact of uncertainties for regulatory decisions lies within the tasks of the decision maker.

3.1.6.2. **EFSA Guidance**

The EFSA guidance document has a specific chapter describing explicitly the roles of assessors and decision makers. It states very similar principles from the Codex Working Principles for Risk Analysis (Codex, 2015):

- "Constraints, uncertainties and assumptions having an impact on the risk assessment should be explicitly considered at each step in the risk assessment and documented in a transparent manner."
- "Responsibility for resolving the impact of uncertainty on the risk management decision lies with the risk manager, not the risk assessors."

It expands on the implications for uncertainty analysis:

"Thus, in general, assessors are responsible for characterising uncertainty and decision-makers are responsible for resolving the impact of uncertainty on decisions" (p.23).

It follows, that the tasks of identification and assessment of uncertainties are to be handled by risk assessors while the resolution is performed by the decision makers. Additionally the guideline recommends increased interaction between the two groups in order to strengthen the mutual understanding of the question at hand and the needs of either group.

3.1.6.3. Summary

The roles of risk assessor and decision makers are the same in both documents. This is not surprising, considering that both institutions have the tasks of assessing risks and not managing them. The EFSA guideline also includes considerations about the interaction by the two groups which aim to help to minimize misinterpretations.

3.1.7. Timing of Uncertainty Analysis

An aspect of uncertainty analysis is when it is supposed to be performed. In principle, this could be during or after the 'original' risk assessment.

3.1.7.1. BfR Guideline

The BfR guideline recommends that uncertainty analysis should be performed parallel to the 'original' exposure assessment. The procedure of uncertainty analysis "should, where possible, accompany the entire process of exposure assessment" (p.9). When describing different tiers of exposure and uncertainty analysis it is additionally stated that "uncertainty analysis must accompany the entire process of exposure assessment" (p.17).

This allows to decide which level of refinement is needed for the exposure assessment since an "accompanying uncertainty analysis might justify the decision at which step the assessment is evaluated as sufficient" (p.15). It also gives an example where uncertainty analysis might impact the original assessment: "[…] the identification of high uncertainty due to neglect of difference and variability might result in the choice of a more sophisticated exposure model" (p.23).

3.1.7.2. EFSA Guideline

The guideline is not explicit in when uncertainty analysis should be performed in relation to the 'original' assessment because it considers uncertainty analysis as a part of it. This can be concluded from statements like this one from planning the assessment strategy: "EFSA's general approach to scientific assessments begins with planning the assessment strategy […] This section expands on aspects of the planning aspects that relate especially to uncertainty analysis. These and the other planning steps may need to be conducted iteratively to arrive at an agreed strategy before starting the assessment, and may need to be revisited and refined later in the assessment process" (p.50). This is part of the wider process of planning the strategy for the assessment as a whole where this part of uncertainty analysis is considered to be part of the overall assessment strategy. The draft EFSA guidance (EFSA, 2016) included a diagram showcasing the general framework for uncertainty analysis, where there was also a distinct pathway where the assessment can be refined based on the result of the uncertainty analysis (see Figure 1).

A meaningful distinction in timing between the two is not possible. As they are considered the same assessment one might conclude that they are to be performed in parallel.

3.1.7.3. Summary

Even, if the BfR guideline is more explicit, both guidelines agree that uncertainty analysis should be carried out in parallel to the other parts of the assessment. They both describe the possibility that uncertainty analysis directly impacts on the risk assessment and therefore trigger a further refinement of the assessment.

3.1.8. Distinction between qualitative and quantitative Uncertainty Analysis

The concepts of qualitative and quantitative uncertainty analysis as well as their specific advantages and disadvantages have been debated extensively in the past. In a first step the definitions of the concepts are compared.

3.1.8.1. **BfR Guideline**

The guideline has a short section on quantitative and qualitative techniques in uncertainty analysis. There the aim of qualitative uncertainty analysis is given as follows:

"Qualitative uncertainty analysis delivers a systematic and comprehensive listing of all sources of uncertainty as well as a discussion of the direction and strength of the influence of uncertainty on the target variable" (p.18).

In the introduction it is additionally stated, that "Qualitative uncertainty analysis aims to create a systematic procedure for the verbal description of inherent uncertainties" (p.10). The guideline also proposes an ordinal scale for description.

Furthermore, the guideline discusses that "Quantitative uncertainty analysis permits the specification of a range of probable values for the target variable (together with numerical ranges for probability bounds)" (p.18). It may also allow "the generation of a probability distribution of the resulting exposure values" (p.18). The expression of uncertainty in a numerical way, be it ranges, bounds or distributions, are considered to be the result of quantitative uncertainty analysis.

3.1.8.2. **EFSA Guideline**

The guideline contains a full chapter on qualitative and quantitative approaches to uncertainty. It considers them as different ways of expression of uncertainty, which consists of two components: "expression of the range of possible outcomes (or a range of values, for a continuous variable), and some expression of the probabilities of the different outcomes" (p.27).

The way qualitative methods express these is "using words, categories or labels. They may provide information on the order of the alternative outcomes, and are sometimes given numeric labels, but they do not quantify the magnitude of differences between the possible outcomes, or their probabilities" (p.27). The guideline furthermore distinguishes between two different types of expression, descriptive expression and ordinal scales. The difference between them is that a descriptive expression is fully verbal while an ordinal scale has ordered categories. Both of them share that they have no way of quantifying the uncertainty. Verbal expressions (either directly or on a verbal scale) are considered to be the result of qualitative uncertainty analysis.

Quantitative methods use numerical scales to describe at least one of the above mentioned components of expression (cp. p.27). Furthermore, "a complete quantitative expression of uncertainty would specify all the outcomes that are considered possible and probabilities for them all" (p.27). As with the quantitative ways of expression the guideline also considers multiple types of quantitative expression of uncertainty:

- **Individual values:** Different values for different possibilities or outcomes
- **Bound:** Either an upper or a lower limit on possible values
- **Range:** Upper and lower limit of possible values
- **Bound/Range with probability:** As above but with a probability
- **Distribution:** Specification of all possible values with a probability

The expression on a numerical scale (possibly with a probability) is thus the result of quantitative methods.

3.1.8.3. Summary

In general, there is an overall agreement on the formal differences between qualitative and quantitative uncertainty analysis. Qualitative methods aim at describing the uncertainties using words or descriptive scales, or more in general verbal methods without a strict mathematical description of the assessed uncertainty. However, when using an ordinal (relative) scale, the uncertainties can still be ordered by strength (regarding its effect on target response)).

Quantitative methods use the language of mathematics to characterise the magnitude of uncertainty and/or probability of different outcomes.

3.1.9. (Dis-)Advantages of qualitative and quantitative Uncertainty Analysis

In a second step, the descriptions of advantages and disadvantages of the two types of uncertainty analysis are compared.

3.1.9.1. BfR Guideline

The BfR guideline does not discuss the advantages of either type in much detail. It considers quantitative methods a higher tier than qualitative methods, and therefore mainly as a refinement of an existing qualitative uncertainty analysis. As such it requires additional and also typically more effort: "If a quantitative assessment is required, in general, it is necessary to get involved scientists with methodological and statistical competence" (p.38f). An advantage of a qualitative assessment would be in that case an easier use and the requirement of less effort. It also discusses that in some cases the quantification of uncertainty might be difficult, e.g. when summarised data or expert judgement are used (cp. p.18), but does not consider quantification impossible in that case.

Basically, the (dis)advantages of qualitative and quantitative expressions are evaluated regarding the degree of refinement of the uncertainty analysis and the amount of work needed for carrying them out.

3.1.9.2. EFSA Guideline

EFSA's guideline clearly encourages the use of quantitative uncertainty analysis. These recommendations are mainly located in section 4.2 (Advantages of quantitative expression) but can be found at many places in the document. In that section, the main issues the guidelines considers in qualitative expression can be summarised as follows (p.28f):

- Because qualitative expressions are ambiguous they may be misinterpreted by decision makers and other stakeholders.
- In some cases the result of an assessment is quantitative (e.g. an exposure of a population). In this case a qualitative expression of uncertainty may not answer the question on how strong the uncertainty influences the result.
- Qualitative expressions may imply indications for how uncertainty influences decision making. Additionally, a qualitative expression will always have to be interpreted by the decision maker which usually understands the sources of uncertainty less than the risk assessors.
- Quantitative uncertainties are more easily combined, namely by calculation opposed to qualitative expressions (e.g. do two mediums become one high or one medium? What about three?).
- Qualitative expressions always contain some sort of subjectivity which, combined with the ambiguity may lead to two different assessors interpreting the same statement differently.

In general, it considers a qualitative assessment of uncertainty to be less preferable than a quantitative one for expressing the overall uncertainty, imposing difficulties in decision making.

Still, the guideline considers that a quantification of uncertainty may not be always possible (p.29f):

- There may be too many uncertainties in an assessment that make the quantification of all of them impractical.
- Quantification of uncertainty may introduce additional uncertainties.
- Or the assessors may simply feel unable to quantify an uncertainty.

In summary, the guideline considers a quantitative expression of uncertainty superior to a qualitative expression but also acknowledges that quantification may not always be possible.

3.1.9.3. **Summary**

The BfR guideline does not explicitly consider major advantages or disadvantages of either way of expression, except for the amount of work needed to carry them out. The EFSA guideline additionally considers the ambiguity and its implications of qualitative expressions a major disadvantage of them. Such a possible ambiguity of qualitative uncertainty analysis is not explicitly addressed in the BfR guideline. Both guidelines still consider quantitative expressions to be more informative than qualitative ones.

3.1.10. Recommendation for qualitative or quantitative Uncertainty Analysis

Finally, the recommendation each guideline gives for choosing a type of analysis are compared.

3.1.10.1. **BfR Guideline**

In its abstract, the guideline states that "Uncertainty analysis is often confined to the qualitative tier" (p.44). It does not give a reasoning why this is the case but considering the previous section this is most likely due to time and resource constraints. Also, as it considers quantitative uncertainty analysis a higher tier, a qualitative uncertainty analysis may often be sufficient for achieving its aims. This is supported by another statement from the abstract: "If necessary, a quantitative assessment of the population exposure and inherent uncertainties is conducted" (p.44).

In certain aspects it also recommends quantitative methods, e.g. "Quantitative sensitivity analyses are strongly recommended/required to describe the potential effect of uncertainties on the final outcome of the assessment" (p.39).

Thus the recommendation of the guideline is mostly implied: It prefers qualitative assessment of uncertainties as long as it is sufficient for the assessment question at hand. Quantitative methods may also be used and are in some cases required.

3.1.10.2. **EFSA Guideline**

The EFSA guideline strongly prefers quantitative methods for expressing the overall uncertainty: "The Scientific Committee concludes that assessors should express in quantitative terms the combined effect of as many as possible of the identified sources of uncertainty" (p.29). This mainly follows from the previously discussed ambiguity of qualitative assessments.

In cases, where quantification is not possible, qualitative methods may still be used. The guideline also points out, that qualitative methods still have their place in uncertainty assessment:

"The Scientific Committee emphasises that qualitative methods are important not only for describing those sources of uncertainty that the assessors are unable to include in the quantitative assessment, but also for prioritising which sources of uncertainty to quantify individually, and for facilitating judgements about the combined effect of sources of uncertainty that are quantified collectively" (p.30).

3.1.10.3. **Summary**

The BfR guideline recommends typically the use of one qualitative method first (due to the comparatively less effort it requires) and the use of quantitative methods later on if deemed necessary. Since the EFSA guidance document requires an overall quantification, assessing the individual uncertainties with quantitative methods (if possible) is preferred.

Still, both guidelines include the other way of expressing uncertainty. In the context of the BfR guideline one could conclude, that a quantification of uncertainty is necessary (the guideline does not describe in a specific way when this is exactly needed). On the other hand, when quantification is impossible the EFSA guideline even recommends a verbal expression or other qualitative method for uncertainty analysis (in contrast to not addressing it at all). It is even advocated to describe them alongside with the results of the quantified assessment.

3.1.11. General Structure of Uncertainty Analysis

The general structure covers all individual steps to be performed for an uncertainty analysis.

3.1.11.1. **BfR Guideline**

The structure of an uncertainty analysis proposed by the guideline follows the steps of the exposure assessment process which are given as follows (p.23):

1. Formulation of the goals and questions of the exposure assessment
2. Exposure scenario
3. Exposure model
4. Parameters of the exposure model
5. Method for exposure calculation
6. Presentation of the findings of an uncertainty analysis
7. Evaluation, interpretation and communication of uncertainties

Steps 1 to 5 address the identification and support the assessment of uncertainties in the corresponding part of the exposure process. Step 6 contains the assessment of the uncertainties as well as the documentation of them. In the final step the guideline considers various points on the communication of the contents of an uncertainty analysis to various audiences.

3.1.11.2. **EFSA Guideline**

The draft EFSA guideline considers the following individual steps for an uncertainty analysis (p.45):

1. Initial plan for assessment strategy
2. Identify and list uncertainties affecting the assessment
3. Select which uncertainties to assess individually
4. Assess individual sources of uncertainty
5. Quantify the combined uncertainty
6. Investigate the influence
7. Describe the unquantified uncertainties
8. Document and report the assessment, including the uncertainty analysis

After a plan phase for the assessment strategy, step 2 considers the identification of uncertainties. The assessment of the uncertainties is covered by steps 3 to 5. Step 6 is an assessment of the impact of individual uncertainties on the result. Finally steps 7 and 8 consider the documentation of either unquantified uncertainties and of the whole assessment. Step 6 is optional in cases where the combined uncertainty is too small to make an impact on the decision-making (cp. p.47). The structure of an uncertainty analysis (including further steps of refinement which are discussed below) is summarised in Figure 1.

The EFSA guideline also introduces a distinction between three different types of assessments which might occur during EFSA's work (p.42f):

- **Standardised procedures:** Assessments where in some way elements which address uncertainty are already included, e.g. by using uncertainty factors or conservative assumptions.
- **Case-specific assessments:** Assessments without standardised procedures or where the existing procedure does not cover all identified or possible uncertainties.
- **Emergency assessments:** Assessments with very little available time and resources.

In assessments with standardised procedures or emergency situations parts of these steps can be omitted either because they are covered by the procedure or due to the time and resource constraints of the situation. This especially concerns the evaluation of individual sources of uncertainties.

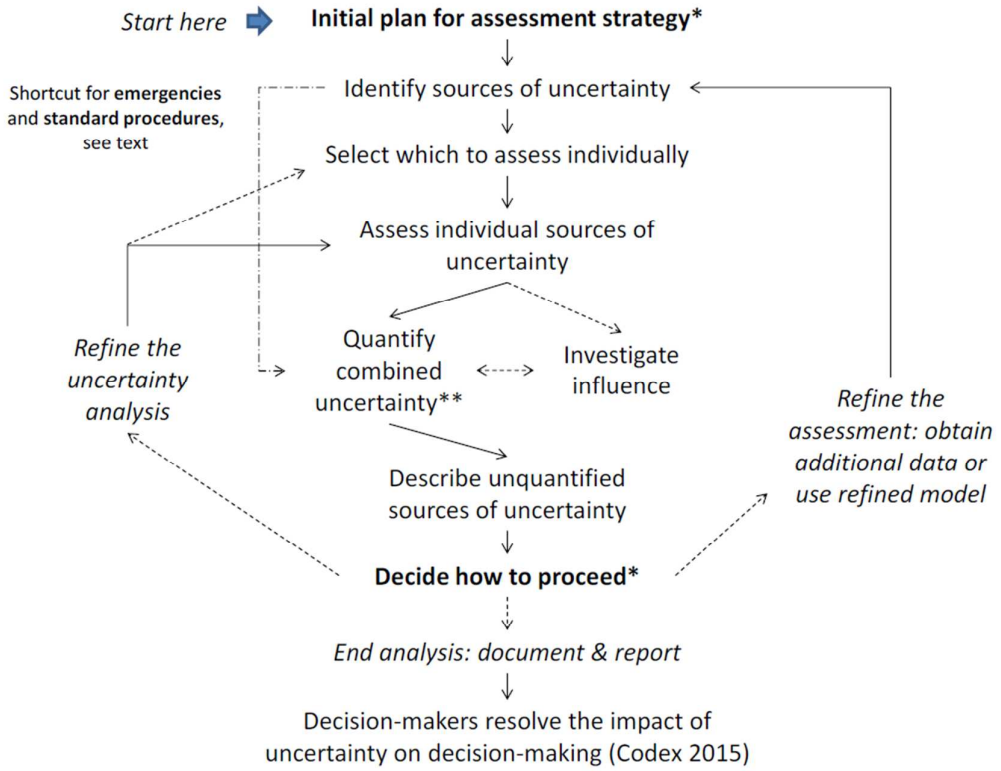

Figure 1: General structure of UA in EFSAs draft guideline (EFSA, 2016)

3.1.11.3. Summary

EFSA's guideline on uncertainty analysis considers more steps than BfR's, especially considering the concept of classifying the sources of uncertainties regarding their assessment (individual or combined) and regarding the resulting uncertainty of all uncertainties combined. These steps in such a direct form are not included within the BfR guideline methodology. The BfR structure of uncertainty assessment reflects more the identification of sources of uncertainties along the process of the initial exposure assessment (from the exposure scenario to model formulation and parameter selection), but no clear distinction between individually and collectively assessed uncertainties is mentioned nor how they are actually combined for the final uncertainty assessment. Still, the first five steps consider more strictly the identification of uncertainty and the latter steps the assessment, documentation and communication.

Additionally, as EFSA's guideline prefers to quantify most uncertainties, it necessarily has an additional step of describing the unquantified sources of uncertainty[3]. Also, the EFSA guideline considers an additional step of planning the assessment.

[3] This is also done in BfR's guideline but not considered explicitly a separate step.

The communication of the uncertainty results is considered part of BfR's guideline in opposite to EFSA. This is a conscious choice as there is a separate handbook on risk communication developed by EFSA.

Still, the major steps which need to be performed in an uncertainty analysis (identification, assessment and documentation of uncertainties, as well as steps for refinement) are included in both guidelines. In the identification of uncertainties the BfR guideline follows more closely the steps of an exposure assessment because of its scope. Both ways of structuring the uncertainty analysis follow a clear logic and are complementary.

3.1.12. Identification of Uncertainties

Identification of the sources of uncertainty is obviously a crucial step in uncertainty analysis. This section compares the approaches of both guidelines.

3.1.12.1. **BfR Guideline**

The guidance provides a set of tables containing detailed questions on the exposure assessment at hand. For each of the first five steps described in section 3.1.11.1 (formulation of the goals and questions of the exposure assessment, exposure scenario, exposure model and method for exposure calculation) a table is provided. An example for such a question is (from table 5 concerning the exposure scenario):

"Are the mechanisms by which a reduction of the concentration/amount of the noxious agent in the contact medium is possible (e.g. air exchange rate, mixing, degradation, decomposition) known and characterised?" (p.28)

Some questions like the above example are rather broad, whilst many of them are very specific, e.g.:

"What are the possible consequences of the inclusion or exclusion of values below the detection or quantification limit?" (p.33).

Table 1 shows the different types/sources of uncertainties the guideline considers. As the scope of the document is focussed on exposure assessment, the questions are related to this topic. Nevertheless, many of them still are applicable to more general assessments, e.g. many questions concerning the exposure model equation can also be applied to more general models. Each of the tables is supported by an own section where the aim and contents of them are explained.

Uncertainty itself is considered in three dimensions which are quoted from (IPCS & IOMC, 2008):

- The level of uncertainty of the exposure assessment.
- Appraisal of the knowledge base of the exposure assessment.
- Subjectivity of choices inherent in an exposure assessment.

For each of these three dimensions an additional list of questions is quoted from the same document which further assists the identification of uncertainty.

Table 1: Types of Uncertainty considered by BfR's guideline

Goal and question formulation of the exposure assessment	Exposure scenario	Exposure model	Parameters	Method for exposure calculation
Question formulation	Development	Estimation of exposure : definition of the target variable	Expert opinions, default assumptions	Deviations
Context	Release/sources	Concept and assumptions for transfer of the scenario into mathematical model	Definition and quantification of the influencing variables	Review of calculations
Population group to be protected	Spread	Connections/ Correlations	Reliability of measurements	Deficient report compilation
Protection goals	Reduction	Model structure, e.g. stratifications	Quality of data sources	Verification
Protection level	Contact: exposed population	Choice of model equation	Study population	
Scope and limitations	Exposure events	Extrapolations of the model	Representativeness	
	Spatial, time-based and situational differences	Risk management measures	Details of correlations/dependencies	
	Risk management measures (RMMs)		Evaluation methodology	

3.1.12.2. **EFSA Guideline**

Similarly, the EFSA guideline also provides tables for the identification of uncertainty. One of them considers assessment inputs while the other one handles the assessments structure (e.g. the used model if applicable). Most of the questions themselves are rather broad, e.g. a question considering the evidence for the assessment structure:

"What is the nature, quantity, relevance, reliability and quality of data or evidence available to support the assumption or judgement?" (p.57).

The aim of these tables is to allow assessors to identify all possible sources of uncertainty, which explains the broad scope of the questions. (cp. p.54). Nevertheless, some of the questions are also specific, e.g.

"How many experts contributed to developing the structure of the assessment or model, how relevant and extensive was their expertise and experience for making it, and to what extent did they agree?" (p.57).

The guideline emphasises that the tables should be applied to each component of the assessment. As the activities of ECHA contain many different tasks the guideline does not give a list of possible components:

"Therefore, this guidance does not offer a general classification of components, but rather recommends that each area of EFSA should consider establishing a list of components for the main types of assessment done in their area" (p.53).

Furthermore, a structured way for identifying all possible uncertainties in the assessment is provided (p.54):

1. Listing of all sub-questions of the main question (e.g. exposure assessment).
2. Listing of all inputs for all identified questions and sub-questions.
3. Identifying all possible uncertainties for each of the inputs for each type of the corresponding table.
4. Identifying all possible uncertainties of the assessment structure for each sub-question using each type of the corresponding table.

The possible types the guideline considers for assessment inputs and structure are summarised in Table 2.

Table 2: Types of Uncertainty considered by EFSA's draft guideline

Assessment inputs	Assessment structure
Ambiguity	Ambiguity
Methodological quality of data sources	Excluded Factors
Sampling uncertainty	Use of fixed values
Assumptions and expert judgements	Relationship between components
Extrapolation uncertainty	Evidence for the structure of assessment
Distribution choice	Calibration or validation with independent data
Other uncertainties	Dependency between sources of uncertainty
	Other uncertainties

Importantly, the EFSA guideline explicitly states that for identifying sources of uncertainty the series of questions used in the BfR guideline (or any other typology) might be adopted instead of the EFSA guideline, if they are considered more suitable for the relevant field (p.54).

3.1.12.3. **Summary**

Both guidelines use a structured way of identifying uncertainties using tables of questions. The questions of the BfR guideline are tailored to its scope of exposure assessment while the EFSA guidelines questions tend to be broader. There are also many more (~100) questions in the BfR list than in EFSA's (~50). This is most likely due to the narrower scope of BfR's guideline; it allows taking knowledge about exposure assessments and their typical uncertainties into account. On the other hand, to take into account all possible situations in its wider scope some questions must be formulated broader in EFSA's guideline. The EFSA guideline acknowledges this fact and proposes that the approach used by the BfR guideline can be adopted if deemed favourably.

Therefore, both guidelines share the same principal understanding regarding the identification of sources of uncertainty in exposure assessment.

3.1.13. Assessment of Uncertainties

Besides the identification of uncertainties their assessment is naturally one of the major tasks of an uncertainty analysis.

3.1.13.1. **BfR Guideline**

The guideline recommends one major way of assessing uncertainties. It is a structured way of documentation of the previously identified uncertainties: "It is recommended to document all identified answers in a structured manner" (p.36). This progress can be supported alongside the identification of uncertainty, e.g. by taking notes which at this stage can have a less standardised form (cp. p. 36). However in the end a "systematic presentation in the form of a table" (p.36) is recommended for assessing them and also starts the documentation process.

In the assessment the same three dimensions as in the identification should be considered and assessed separately:

- Degree of uncertainty
- Confidence in the knowledge base
- Subjectivity of choices

All of them should be assessed for each of the different types of uncertainties summarised in Table 1.

A scale is proposed for judging the degree of uncertainty containing and is shown in Table 3. It is proposed with "the aim […] to highlight the primary sources of uncertainty" (p.37) affecting the exposure assessment.

Additionally to this qualitative method of assessment the guideline also considers quantitative methods of uncertainty estimation and references (IPCS & IOMC, 2008) for further details and states that these may be categorised as (p.35):

- the calculation of the lower and upper limits of the exposure in the form of an interval estimate
- the application of probabilistic (distribution-based) methods
- sensitivity analysis

Sensitivity analysis is the only of those methods for which some details are given. The guideline contains a section on sensitivity analysis, where it is described in the following way:

"For the purpose of comparison all parameters might be changed by a fixed rate (e.g. ± 20 %), by changing the parameter values by one unit (e.g. number of product applications per day) or by empirical ranges (e.g. using mean ± standard deviation intervals or by applying uncertainty

distributions. Mathematical methods for the comparison of results range from direct numerical comparison, evaluation of rates for change up to more complex regression techniques. First-tier methods generally use a calculation technique in which only one parameter of the model is changed at a time relative to a standard case (e.g. a mean/median value for all other parameters). In distribution-based (probabilistic) sensitivity analysis, the degree of influence of the relevant influencing factors can be simultaneously quantified for multiple variables with the help of statistical methods" (p.19).

It does not give further instructions on how to perform a sensitivity analysis but continues on discussing further applications for it, e.g. in the development of a model where it might indicate factors where more precision (because of high sensitivity) is needed in the modelling as others (cp. p.19).

Table 3: Ordinal scale as proposed by BfR's guideline for assessing uncertainty (p. 37)

Degree of potential effect	Possible direction		
	Underestimation	Not known/Underestimation and overestimation possible	Overestimation
Not discernible/ Negligible	0: Uncertainty has no discernible or a negligible effect on estimation of the risk	0: Uncertainty has no discernible or a negligible effect on estimation of the risk	0: Uncertainty has no discernible or a negligible effect on estimation of the risk
Low	–: Uncertainty can result in a low underestimation of the risk	–/+: Uncertainty can result in a low deviation in the estimation of the risk in both directions	+: Uncertainty can result in a low overestimation of the risk
Moderate	– –: Uncertainty can result in a moderate underestimation of the risk	– –/++: Uncertainty can result in a moderate deviation in the estimation of the risk in both directions	++: Uncertainty can result in a moderate overestimation of the risk
High	– – –: Uncertainty can result in a high underestimation of the risk	– – –/+++: Uncertainty can result in a high deviation in the estimation of the risk in both directions	+++: Uncertainty can result in a high overestimation of the risk
Not known	? –: Uncertainty can result in a underestimation of the risk of unknown magnitude	? –/+: Uncertainty can result in a deviation in the estimation of the risk in both directions and of unknown magnitude	? +: Uncertainty can result in a overestimation of the risk of unknown magnitude

Furthermore, it recommends a stepwise evaluation of the scenario uncertainty. In case of dependencies, i.e. plausible or empirically observed correlations among model input parameters, it usually considers qualitative expression as sufficient but if a parameter shows strong influence or has a strong association with another parameter of the model then a quantitative approach should be used (cp. p. 35).

3.1.13.2. **EFSA Guideline**

Since one of the aims of the guideline is to provide a toolbox of various methods for performing uncertainty analysis, a long list of methods is supplied:

- Expert discussion
- Expert Knowledge Elicitation (EKE)
- Descriptive expression
- Ordinal scales
- Matrices
- NUSAP
- Uncertainty table for quantitative questions
- Uncertainty table for categorical questions
- Interval Analysis
- Confidence Intervals
- The Bootstrap
- Bayesian Inference
- Probability Bounds Analysis
- Monte Carlo
- Approximate probability calculations
- Conservative assumptions
- Sensitivity Analysis

Each method has a short description in the text and a longer description in the annex. It is further supported by an overview table summarizing the capabilities of each method as well as important information regarding selection criteria using the following categories:

- Evidence of current acceptance
- Expertise needed to conduct
- Time needed
- Theoretical basis
- Degree/Extent of subjectivity
- Method of propagation
- Treatment of uncertainty and variability
- Meaning of output
- Transparency and reproducibility
- Ease of understanding for non-specialist

The guideline also has a section on explaining how a suitable method can be selected for each uncertainty. It expands on these categories by also considering more practical aspects, e.g. that initially a method similar to the original assessment can be used for uncertainty assessment. It also contains a step-by-step guide for the selection process. The guideline also allows the use of further

methods if they are suitable: "Assessors are free to consider other methods that they consider suitable" (p.85).

As most of the methods are not contained in BfR's guideline at all, only a few of them will be discussed in a bit more detail. These are "Uncertainty Tables for quantitative questions" and "Sensitivity Analysis".

Uncertainty Tables for quantitative questions

The guideline describes uncertainty tables for quantitative questions as a way of a structured listing of all identified uncertainties as well as a description of their individual and combined impact. This is presented in a table with two or more columns. Direction (and sometimes strength) is depicted by using + and – signs (multiple ones for strength). These scales may remain ordinal but can also be quantified by assigning a numerical value to each symbol or combination of symbols (cp. p. 158).

The guideline considers this method applicable to "all types of uncertainty affecting quantitative questions or estimates, in all areas of scientific assessment" (p.162). It also positively mentions its structured way of reporting uncertainty as well that it "makes transparent many subjective judgements that are unavoidably present in risk assessment" (p.161). However, the qualitative way of expressing uncertainty using symbols may "be prone to misunderstanding" (p.162), even when a quantitative scale is used alongside with them. Also, it mentions that the method extensively uses "expert judgement, which is subject to various psychological biases" (p.162). The guideline would prefer the impacts represented by a numerical scale and the use of formal elicitation methods to combat these issues (cp. p. 162).

Sensitivity Analysis

The guideline describes multiple ways of performing sensitivity analysis, grouped into three categories: Graphical, deterministic and probabilistic. For each of the methods it also lists a couple of software packages capable of performing them. It considers graphical methods as a complement to the other methods, visualizing their results (cp. p. 253) and thus does just shows an example for all of them without discussing them in much detail. For each of the other methods a short summary describing their main concept is given.

The 'simplest' method described in the guideline 'Nominal Range Sensitivity Analysis' in which a models input are changed around its nominal values by some bound. It however fails to take interactions between multiple inputs into account. Furthermore, different types of regression and probabilistic methods are described in the guideline.

In general, sensitivity analysis is regarded as a "valuable complement of uncertainty assessment in EFSA" (p.267), because "it helps assessors in providing risk managers with information about most influential factors on which to focus actions and further research" (p.267). A downside is that some of the methods "need to involve an experienced statistician in the computation and interpretation" (p.267). It also points out, that it "is necessary to clarify prior to start the sensitivity analysis which question it is intended to reply" (p.267) in order to not limit its value.

3.1.13.3. **Summary**

EFSA's guideline contains many more methods for assessing uncertainties. This is the result of different aims of the two guidelines, since EFSA's document wants to provide a toolbox whereas BfR's aim is simply to provide framework in which uncertainty analysis can be performed.

The methods described by BfR are contained in EFSA's toolbox – the tabular form of assessing uncertainty is an example of an uncertainty table, albeit EFSA's guideline would prefer a numerical scale associated with the method instead a purely qualitative one. Sensitivity analysis is also recommended by BfR's document without providing much detail. This detail is provided by EFSA's

document in its description of sensitivity analysis. BfR's guideline also allows other ways of assessing uncertainty but does not describe them. Any method described in EFSA's guideline would thus also be usable by an uncertainty assessment according to BfR's guideline.

3.1.14. Additional steps in the EFSA guideline

There are a couple of additional steps in performing an uncertainty analysis within EFSA's guideline. In this section they are summarised and their implications for the project are discussed. Some of these topics are also touched by the BfR document and will be mentioned in the corresponding place.

Similarly, the BfR guideline discusses communication of uncertainty which is not part of EFSA's guideline. This will be discussed in a later section.

3.1.14.1. **Planning of the Assessment Strategy**

Before actually starting the assessment, EFSA's guideline requires a step in which the assessment strategy is planned. This is part of the larger strategy for the whole assessment which contains several steps:

"Clarifying the scope of the assessment, developing the conceptual framework for the assessment, defining the evidence needs, and planning the approaches to be used for collecting data, for appraising evidence, for eliciting expert knowledge, and for analysing and integrating evidence" (p.50).

These steps may be needed to be performed iteratively until an agreed strategy is reached. The section itself discusses two steps particularly important for uncertainty analysis: The definition of the question and the plan for the uncertainty assessment itself.

Definition of the question

In this step, the question to be addressed by the assessment needs to be checked. Ideally, it should contain no ambiguity and not imply impacts for risk management. It is recommended that the terms of reference are checked word for word and potentially ambiguous words are either defined or replaced by unambiguous ones. This may also require a direct discussion with the risk manager for clarification (cp. p. 50).

The aim of this procedure is to reduce potential uncertainties introduced by an imprecise question and to ensure that the result of the assessment can be interpreted in the way it was intended by the risk manager.

The aspect of potential uncertainties from the research question is also considered in BfR's guideline. There, one of the tables for identifying uncertainty specifically is for uncertainties originating from the task definition (Table 4, p.26). It is also the first table to be used and the corresponding section includes similar recommendation for further clarification if necessary:

"Before beginning with the assessment, uncertainty analysis of the goals and questions should reduce the ambiguity of the scope question, it should pinpoint aspects that are unclear and identify alternative interpretations. This helps to clarify the situation and supports effectivity" (p.25).

Planning the uncertainty analysis

In the planning of the uncertainty many key decisions for the uncertainty analysis are taken. They include splitting the assessment into sub-questions if needed, the decision if the assessment will be quantitative or qualitative as well as the development of a conceptual model. Further important points are an assessment of time and resource constraints and the availability of knowledge and/or data gaps (cp. p.52).

The guideline also states that some of the planning steps may be skipped in case of standardised or emergency assessments.

3.1.14.2. Individually and combined Assessment of Uncertainties

EFSA's guideline recommends that the most important uncertainties should be assessed individually: "As the sources of uncertainty that are included in the initial subset will receive more detailed consideration, it makes sense to prioritise them based on the potential magnitude of their impacts on the uncertainty of the assessment outcome or conclusion" (p.83). Each of these uncertainties is then assessed individually and their impact on the component it affects is estimated, either qualitatively or quantitatively. The remaining uncertainties are treated collectively.

As the step of selecting an initial set for individual assessment is rather early in the assessment strategy, a pragmatic way is recommended by "considering each uncertainty briefly in turn and prioritising them by expert discussion or semi-formal expert elicitation of:

- their potential impact on the assessment outcome (see preceding paragraph), and
- the availability of data and readily-applicable methods to assess them" (p.84).

At a later stage in the analysis more uncertainties may be assessed individually (which have been assessed collectively before) if deemed necessary to answer the assessment question.

3.1.14.3. Quantification of the overall Uncertainty

The quantification of combined uncertainty has three aspects in the guideline:

- Combining the effects of the individually assessed uncertainties using a quantitative method.
- Combined assessment of all uncertainties which were either assessed qualitatively or not individually ('Not quantified individually').
- Combination of the results of the above steps.

For the first part the guidance lists various methods depending on the way the individual uncertainties were quantified (e.g. probabilistically and deterministically or the use of confidence intervals for one uncertainty). If the same method has been used for multiple uncertainties a corresponding form of calculation can be used (cp. p. 93).

The second step starts by judging if the uncertainties which have not been quantified individually will make a significant contribution compared to the quantified uncertainties. If this is considered not to be the case it may be concluded that the total quantified uncertainty covers their impact as well. Otherwise expert judgement should be used to quantify their combined impact (cp. p. 94).

Ideally, the resulting uncertainty is expressed quantitatively and the methods recommended in the first step may be used for a final combination. Potential dependencies between the uncertainties should be taken into account. The guideline also mentions situations where this may not be possible, in which case expert judgement should be used for combination (cp. p. 94).

3.1.14.4. Summary

Regarding the planning of the assessment, many of the points listed in EFSA's guideline are touched by BfR's guideline. However, they are often not described in explicit steps. Instead, they are considered as potential sources of uncertainty (e.g. data gaps) or implicitly (the need to use another method after an initial assessment). At this point it is worthwhile to point out that also the EFSA guideline considers refinement of the assessment strategy if needed. In case of the definition of the assessment question many of the points considered by EFSA's document are part of the first step of

identifying uncertainty in BfR's guideline. As BfR's guideline uses uncertainty tables for assessment of all identified uncertainties it does not have the concept of combining different sources of uncertainty.

As has been pointed out before, BfR's guideline encourages the use of other methods than the one described in more detail. It contains little information how to proceed if the results need to be combined.

3.1.15. Refinement of Uncertainty Analysis

This section compares the conditions for a refinement of uncertainty analysis and practical implementations for this refinement.

3.1.15.1. **BfR Guideline**

In the basic principles, which the BfR guideline quotes from the WHO-IPCS document, point 9 reads:

"Where appropriate to an assessment objective, exposure assessments should be iteratively refined over time to incorporate new data, information and methods to better characterize uncertainty and variability" (p.13). This is handled by a multi-tier approach of exposure assessment which is motivated in part by managing the workload of the assessor with regard to the available resources:

"Multi-tier methods are standard practice in the field of exposure assessment. By this, the workload for exposure assessment can be limited to the scope required to assess the achievement of the protection goals" (p.15). Higher tiers typically use more data and more sophisticated methods and thus allow a higher level of precision. Lower tiers typically offset this lack of precision by using conservative default assumptions which should lead to an overestimation of exposure. The guideline groups the exposure assessments into three different tiers (p.16):

- **Initial exposure assessment:** Generic scenarios and default parameters.
- **Deterministic exposure assessment:** Specific scenarios and possibly stratifications. Parameters are obtained from descriptive statistics from data wherever possible.
- **Distribution-based exposure assessment:** Specific and refined scenarios. Full distributions are used for parameters wherever possible.

This concept is later expanded to a multi-tier concept of uncertainty analysis. Again the higher the tier the higher typically the sophistication of used techniques and amount of effort needed to execute. The following three tiers are considered (p.16f.):

- **Application of uncertainty factors:** Multiplying the result of exposure assessment by fixed values to address uncertainties or estimating a "margin of safety" by comparing the exposure to some reference value.
- **Qualitative uncertainty analysis:** Identification and documentation of existing uncertainties.
- **Quantitative uncertainty analysis:** Quantification of uncertainties and estimating their impact on the assessment result.

Each of these tiers can in principle be used with each of the tiers of exposure assessment even though it states that "A lower level of iteration will generally also result in a lower step of the uncertainty analysis" (p.17).

3.1.15.2. **EFSA Guideline**

EFSAs guideline "recommends a flexible, iterative approach, which refines the uncertainty analysis progressively as far as it is needed, rather than a fixed set of tiers" (p.46). It additionally states:

"All aspects of scientific assessment, including uncertainty analysis, should be conducted at a level of scale and complexity that is proportionate to the needs of the problem and within the time and resources agreed with the decision-makers" (p.46). It does not introduce different level of tiers but states that it "distinguishes two main classes of methods for uncertainty analysis, qualitative and quantitative" (p.46).

The three different types of assessments which differ in how the refinement process should be taken into account (p.47ff):

- **Standardised procedures:** The use of standardised procedures should already account for the uncertainties. The analysis should thus first focus on checking if the assumptions in the procedure apply and whether uncovered sources of uncertainty exist. If yes the assessment becomes case-specific.
- **Case-specific assessments:** Procedural refinement as described below. The assessment "should start at a level that is appropriate to the assessment in hand" (p.47).
- **Emergency assessments:** At least the identification of major sources of uncertainty should be performed. Assessment of the impact of individual sources might be omitted and a combined impact should be given. An emergency situation should be followed by a case-specific assessment later on.

A refinement should in general be performed if the result of the assessment is not sufficient to inform decision making (cp. p.47). In general it "should be targeted on those sources of uncertainty where refinement will contribute most efficiently to improving the characterisation of uncertainty" (p.47). In general, refinement means either the identification and allocation of new data or the use of more sophisticated models. Another way of refining the uncertainty analysis is to consider more sources of uncertainty individually. The more uncertainties are assessed (and quantified) individually, the more refined the uncertainty analysis is considered.

3.1.15.3. **Summary**

Both guidelines agree that uncertainty analysis should scale with the needs of the assessment at hand. It suffices using a comparatively low level of refinement, if the answer of the initial assessment question is robust in the light of the identified uncertainties. Both guidelines also agree that certain usage of data and or methods for addressing uncertainty are more sophisticated/refined than others. One clear division line both guidelines use is the distinction between qualitative and quantitative methods.

The guidelines differ in naming the different levels of sophistication. The BfR guideline explicit introduces 'tiers' of uncertainty analysis while the EFSA guideline just considers different levels of refinement without specifying them. Also, as the EFSA guidelines considers the possibility that different sources of uncertainty are assessed differently and contains the concept of individual/combined assessed uncertainties, it has an additional aspect of refinement: the number of individually assessed uncertainties. It also considers different ways of approaching refinement based on its different types of assessment (standardised, case-specific and emergency).

The differences – except for the naming conventions – result from the different approaches on methodology.

3.1.16. Documentation

Finally, the way each guideline recommends the documentation of the uncertainty assessment and its results is compared.

3.1.16.1. **BfR Guideline**

The guiding principles the guideline uses requires the documentation of uncertainties to be in "a comprehensive, systematic and transparent manner and should take account of qualitative or quantitative aspects of the methods, scenarios, models, parameters, data, findings, sensitivity analysis and interpretations of results" (p.13). In praxis, the documentation of uncertainties is closely related to their assessment because the guideline proposes to use the same table developed for the assessment of uncertainties to be used for documentation (cp. p. 36f).

Additionally, "the uncertainties with the greatest relevance should be summarised" (p.37). The following points should allow identifying the uncertainties with the greatest relevance:

- "Which sources and reasons for inherent uncertainties have high importance?"
- "What are the effects (degree and direction) of the most important identified uncertainties on the result of the exposure assessment? If defined, can the level of protection be warranted in view of the reported uncertainties?"
- "What are the specific options for uncertainty reduction in the exposure assessment? Are these options suitable to allow a sufficient and appropriate assessment of the protection goals?" (p.37)

For this, "sensitivity analysis might substantiate the qualitative results" (p.37).

3.1.16.2. **EFSA Guideline**

The guideline states that "it is essential for transparency to document in a concise and clear way all of the sources of uncertainty identified and how they have been addressed in the assessment" (p.82). The guideline recommends that "every assessment report should include a section on uncertainty analysis" (p.98).

In standardised procedures not much effort is needed to achieve this. A reference to the describing document, where the standardised way of taking uncertainties into account, should be given, and – if needed – a short explanation why that assessment is applicable and sufficient (cp. p.97).

In other assessments the following points should be included in the report (p.98):

- The assessment questions and sub-questions
- A list of identified uncertainties and the way they were prioritized for individual assessment (if applicable)
- For each individual assessed uncertainty the method of assessment
- The method used for combination of uncertainties
- The result of the uncertainty assessment and their relative influence on the assessment result
- A summary table

The draft guideline includes a proposal for such a summary table, although this was not included in the final version (Benford *et al.*, 2018).

Additionally, a "concise summary of the conclusions in format and style suitable for inclusion in the executive summary of the overall assessment report" (p.99) should be prepared which should use language which can be understood by non-experts.

3.1.16.3. **Summary**

Both guidelines stress the importance of transparency in the documentation of uncertainties. Both propose a structured way for the documentation itself. EFSA's guideline is more specific in what exactly needs to be documented but the general requirements are very similar.

A template summary table for the documentation of uncertainties is provided by the BfR document and the draft EFSA guideline, but not the final EFSA guideline. Both BfR and EFSA propose a shorter summary for the report for decision-making, focussing on overall uncertainty and the description of the most important uncertainties.

3.1.17. Communication of Uncertainties

The assessed version of EFSA's guideline does contain a section on communication, which will be removed in the currently ongoing revision. Thus only the viewpoint of BfR's guideline on this topic is stated.

3.1.17.1. **BfR Guideline**

The guideline bases its communication strategy on an EFSA document outlining principle communication aspects (EFSA, European Food Safety Authority, 2012). The following aspects are general criteria introduced by that document:

- Comprehensibility
- Usability
- Transparency
- Up-to-date character

Considering the specific risk at hand, the following information should also be provided:

- Affected groups
- Voluntariness/Controllability
- Severity

The guideline then expands on these general requirements by discussing their implication on uncertainty analysis. The basic principles of this are given as follows:

"The assessor should try to predict which aspects of uncertainty analysis will be of particular relevance for the risk management and for the public at large. The communication of an uncertainty assessment should in particular address who may be affected by the uncertainties, how serious the possible impacts of uncertainties might be, and what can be done to control and reduce these impacts" (p.40).

The guideline also lists a group of questions related to uncertainties which should be answered by their communication (p.40f):

- Which findings can be reported as "based on a sound scientific knowledge", and which ones are to be categorised as "uncertain"? How can this be presented in a way that the abstraction level of the answers corresponds to the knowledge of the population?
- Which are the uncertain elements with the greatest influence on the result of the exposure and risk assessment? Which causes and sources contribute to existing uncertainties?

- Who (which groups) and how the population might be affected by the uncertainties?
- Which assumptions were used to solve an assessment task despite inherent uncertainties? How are these assumptions justified? Which diverging opinions on these assumptions are known to the assessors? The choice of assumptions should be explained.
- Which measures can be taken to reduce or remove uncertainties? How to procure missing information? Which resources (e.g. workload, human and laboratory resources, infrastructure and budget) are necessary for this purpose?
- Which recommendations can be given, e.g. for improving regulation, enforcement of existing rules, contracts and laws, monitoring and preventive measures?

3.2. Summary of the comparison of the respective uncertainty documents

The comparison of both guidelines (BfR and EFSA) leads to the overall conclusion that they both establish a framework for uncertainty analysis, inspired by the same philosophy. Topics considering the motivation, the principal sequence, as well as the documentation of uncertainty analysis are described similarly. However, it should be kept in mind that BfR's guideline is only applicable for uncertainties in exposure assessments in contrast to EFSA's guidance, which is applicable for any risk assessment within EFSA mandate. Therefore, a comparison of both guidelines is only feasible regarding uncertainty analysis in exposure assessments.

Apart from the scope, they are two main differences between the documents. First, EFSA's document describes a broad range of qualitative and quantitative methods for uncertainty analysis, while the BfR guideline explicitly characterises only one (uncertainty tables), which are also described in the EFSA guidance. However, this does not imply that BfR's guideline discourages the use of quantitative methods. In contrast, after applying this approach, risk assessors are free to pursue any quantitative method of their choice. Second, EFSA prescribes that the overall uncertainty is quantified, emphasising that qualitative expressions are easily misinterpreted by decision makers. EFSA recommends that the most important sources of uncertainty are assessed using quantitative methods, since it will make the overall quantification much easier. However, the use of qualitative methods is by no means prohibited. In contrast, the BfR guideline does not require a quantification of the overall uncertainty, nor is in general the use of quantitative methods mandatory. It should be noted, though, that the EFSA guidance does not demand quantification of individual sources of uncertainty either, but requires an overall quantification of uncertainty.

As a result, an uncertainty analysis following the EFSA guidance regarding an exposure assessment will in principle comply with the BfR guideline. Vice versa, the situation is more complicated. An uncertainty analysis without an overall quantification might be very well in accordance with the BfR guideline. However, a missing overall quantification is not in line with the principles of EFSA's guidance document.

The implications of these differences regarding the project are as follows:

- The scope of the uncertainty analysis of the case studies needs to be restricted to exposure.
- In accordance with the EFSA guidance document, the overall uncertainty of both case studies will be quantified. For the BfR guideline, uncertainty tables will be used.

A preliminary draft of the new version of the EFSA guidance will be used for both case studies.

3.3. Rationale for selection of case studies

A couple of examples from the work of BfR were selected as candidates for the two case studies. All of them originate from food risk assessments:

- Cadmium and Lead in infant and child food
- Tropane alkaloids in peppermint tea
- Aluminium in cocoa
- Strontium chloride and Alizarin red S in eel
- Manganese in (general) food
- Contaminants in game

The examples were categorized by their time frame, the level of routine (expressed as 'level of standardisation' and the amount of available data[4]). To test the guidances in different situations, the case studies should vary in all of these aspects in order to allow testing of the guidelines in different situations. A summary of the classification of each of these examples is shown in Table 4.

In order to select examples with a low level of routine only two examples remain applicable: "manganese in food" and "Strontium chloride/alizarin red S in eel". Subsequently, the example concerning eel is the only one which applies for a rather low amount of data. Therefore, this example should be selected to cover a low level of routine and little available data. The available time for this example was also rather limited, being in the order of one to two weeks. Due to the low amount of time and available data this example would also work well as the purely qualitative one.

Two examples remain which differ in all three categories from the one concerning eel: "aluminium in chocolate" and "cadmium and lead in infant and child food". Both of them were routine evaluations using routine data as well as having larger amounts of time (1 month and 1-2 months respectively). The main difference in terms of routine between them is that the example concerning chocolate was an assessment of the whole population (leading to separate assessments for children and adults) while the other one only concerns infants. From that point of view the former example seems to be more interesting for an extended uncertainty analysis.

Thus, the following two examples are proposed:

First case study: Strontium chloride/Alizarin red in eel

Second case study: Aluminium in chocolate

[4] Also by applicable panel but basically all case studies were within scope of CONTAM

Table 4: Categorization of candidate case studies into selection criteria

Example	Panel	Time frame for original assessment	Level of standardization in approach	Available data
Cadmium and Lead in infant and child food	CONTAM	Long (1-2 Months)	High (routine procedure)	Large
Tropane alkaloids in peppermint tea	CONTAM	Short(3 Weeks)	Medium-High (routine procedure, non-routine data source)	Medium
Aluminium in cocoa	CONTAM	Medium(4 Weeks)	High (routine procedure)	Large
Strontium chloride and Alizarin Red S in eel	ANS and/or CONTAM	Very short (1-2 weeks)	Low (Not enough data for routine analysis)	Little
Manganese in food	CONTAM and/or NDA	Long (own analysis)	Low (Use of TDS data and MCRA)	Large

4. First case study – Alizarin red S and strontium chloride in European eel

4.1. Introduction to the first case study

4.1.1. Description of the Case Study and Assessment Question

The European eel (*Anguilla anguilla*) has been listed as a critically endangered species by the IUCN red list of endangered species (Jacoby & Gollock, 2014). The European Union has requested eel management plans from its member states in order to secure the European eel population (Council Regulation (EC) No 1100/2007). In Germany, the most important measure is restocking, where young eels (glass-eels) are caught in estuaries and subsequently released in inland waters. One possibility to determine the efficiency of these stocking measures is the marking of eel with Alizarin red S (ARS) or strontium chloride (SC). The marking is performed by placing eels in a solution containing the marker substances over a time period ranging from several hours up to one day. ARS forms complexes with calcium ions, which can get incorporated into the bones. These incorporations can be studied via fluorescence microscopy. The marking with SC works in a likewise fashion: strontium ions have similar features as calcium ions and are therefore also incorporated into the bones, which can be observed using electron microscopy.

The German Federal Ministry for Food and Agriculture asked the BfR to determine whether the marking of eel poses any threat to consumer health. Since the uncertainty analysis will be carried out for the exposure assessment, the presentation of the consumption data is emphasised in the following.

4.1.2. Summary of the BfR Risk Assessment Report

The BfR report concluded that a risk assessment regarding ARS/SC in eel is not feasible. First and foremost, the toxicity of the substances cannot be rated: For ARS, there is not sufficient reliable data to assess its toxicity, while in case of SC, only the subchronic toxicity can be assessed using the study performed in (Kroes *et al.*, 1977). The no observed adverse effect level (NOAEL) was determined to be 10 mg/kg strontium using rats. However, the genotoxicity, carcinogenic activity, chronic toxicity, and reproductive toxicity among others could not be assessed for strontium. Furthermore, it was stated that there is no available data for ARS/SC concentration in eel tissue. Finally, the existing data regarding eel consumption was listed, which included two 24h recall data surveys from the German National Dietary Survey II (MRI, 2008) and telephone interviews of rarely consumed foods (Ehlscheid *et al.*, 2014). From the two 24h recall data, only 19 participants (0.15%) reported to have eaten eel, on average 35 g per day. In contrast, the telephone survey revealed that 34 % of all participants have eaten eel during a one year period. Altogether, the eel consumption data has been rated as insufficient for performing an exposure assessment but is listed in the following subsection.

4.1.3. Eel consumption data

The uncertainty analysis carried out by both guidance documents will refer to the following data on eel consumption in Germany. These data was available at the time of the initial risk assessment performed by BfR and also shortly summarized there.

The daily eel consumption data were obtained from two 24h recall data surveys as part of the German National Dietary Survey II (MRI, 2008) conducted in 2005-2007. In total, 13926 participants (14-80 years old) were interviewed on two days, questioned about kind and amount of food consumed during the last 24 hours. The participants determined the weight of the food by comparison with a picture book. In total, only 19 incidences of eel consumption were reported. The determined daily consumed portion sizes varied from 22.8 g to 126 g.

The frequency of eel consumption was collected during telephone interviews of rarely consumed foods in 2011 (Ehlscheid *et al.*, 2014). The results are summarized in Table 5. In total, 1004 persons were interviewed (not younger than 14 years old), and 34% reported that they have eaten eel at least once during a one year period. 12 persons (1.2%) claimed to have eaten eel once per week, and one person reported to have eaten eel 2-3 per week.

It should be noted, that these data do not differentiate whether European eel, or other eel species were consumed.

Table 5: Frequency of eel consumption (Ehlscheid *et al.*, 2014)

Frequency	% of consumers
1-5 times per year	22.9%
6-11 times per year	5.1%
1-3 times per month	4.6%
once per week	1.2%
2-3 per week	0.1%
4-6 per week	0
daily	0

4.2. Uncertainty Analysis following the BfR Guideline

4.2.1. Approach

The main characteristic of the original exposure assessment of the BfR is that no exposure value was estimated. This has implications for the uncertainty analysis using the BfR guideline, since no model was built to determine exposure. Therefore the question list regarding the exposure model, the method of exposure calculation as well as the documentation of the exposure calculation is not applicable. Moreover, the subjectivity of choices cannot be evaluated, since no choices were made. Although there is no model, we considered the following three quantities as parameters:

1) ARS/SC concentration in edible eel tissue
2) daily European eel consumption for a consumption incidence
3) frequency of European eel consumption

Subsequently the impact of the uncertainties on the final result – intake of ARS/SC per bodyweight per time period was evaluated. But since no exposure estimation was performed, for some sources of uncertainty only the magnitude of the uncertainties could be assessed, but not its direction.

Moreover, for employing the BfR guideline, the situation prevalent at the time of the original risk assessment (March 2017), especially the available and actually used knowledge, determined the answers to the respective question lists. It is important to stress at this point, that only 1-2 weeks were available for the original exposure assessment, which is a common timeframe for exposure assessments performed at BfR.

The uncertainty analysis starts with an assessment of the knowledgebase and follows with the question lists for assessing the degree of uncertainty. The summarized results of the uncertainty analysis are then presented in a table and the uncertainty analyses ends with the conclusion.

4.2.2. Question lists

The BfR guideline for uncertainty analysis considers all three dimensions of uncertainty:

1) Degree of uncertainty
2) Appraisal of the knowledgebase
3) Subjectivity of choices

Therefore the appraisal of the knowledgebase (Table 7 in section 4.2.2.1) is rated separately from the degree of uncertainty. The rating for this dimension (in low, medium, and high) takes place at the standardised presentation of the findings of uncertainty analysis in Table 11 for relevant topics. The remaining question lists deal with the degree of uncertainty for the definition of the task (Table 8 in section 4.2.2.3), the exposure scenario (Table 9 in section 4.2.2.4), and the model parameters (Table 10 in section 4.2.2.5). The degree of uncertainty is assessed for both, magnitude and direction of the uncertainty. The following symbols used in for assessing the uncertainties defined in Table 6 (BfR guideline p.37 (Heinemeyer *et al.*, 2015).

Table 6: Ordinal scale for classification of uncertainty

Degree of potential effect	Underestimation	Direction not known	Overestimation
Not discernible/ Negligible	0: Uncertainty has no discernible or a negligible effect on estimation of the risk	0: Uncertainty has no discernible or a negligible effect on estimation of the risk	0: Uncertainty has no discernible or a negligible effect on estimation of the risk
Low	-: Uncertainty can result in a low underestimation of the risk	-/+: Uncertainty can result in a low deviation in the estimation of the risk in both directions	+: Uncertainty can result in a low overestimation of the risk
Moderate	--: Uncertainty can result in a moderate underestimation of the risk	--/++: Uncertainty can result in a moderate deviation in the estimation of the risk in both directions	++: Uncertainty can result in a moderate overestimation of the risk
High	---: Uncertainty can result in a high underestimation of the risk	---/+++: Uncertainty can result in a high deviation in the estimation of the risk in both directions	+++: Uncertainty can result in a high overestimation of the risk
Not known	?-: Uncertainty can result in a underestimation of the risk of unknown magnitude	?-/+: Uncertainty can result in a deviation in the estimation of the risk in both directions and of unknown magnitude	?+: Uncertainty can result in a underestimation of the risk of unknown magnitude

A summarized overview of the degree of uncertainty will be given in Table 8 in section 4.2.3.

4.2.2.1. Question list for assessing the knowledge base

Table 7: Question list for assessing the knowledge base

Criteria	Questions[8]	Assessment
Completeness	Was the essential and relevant knowledge base compiled in a manner that is necessary to obtain an exposure assessment with the desired accuracy?	No, the knowledge base was not sufficient to obtain an exposure assessment with the desired accuracy.
	Were the most important deficiencies in the knowledge base identified?	Yes, the most important deficiency is the lack of data for ARS/SR concentration in edible eel tissue. Moreover, there is no eel consumption data for children younger than 14 years. 24 h recalls only yielded 19 cases of eel consumption, which does not allow for any stratification.
	Were the possible effects of these weak points on the result of the exposure assessment controlled?	No exposure was estimated.
	Were assumptions identified that can compensate the weak points of the knowledge base?	No exposure was estimated.
	Were all parameter values and results controlled using comparative calculations?	Not applicable
	Were all dependencies and interrelationships between model variables reviewed?	No
Reliability	Was the knowledge base checked for factual and methodological justifications?	Yes, regarding eel consumption, the consumption survey used is a scientifically respected approach as well as the other data source (telephone interviews).
	Was the knowledge base reviewed to ensure it is scientifically up-to-datedness?	The data from the consumption surveys are from 2005-2007 and 2011, respectively.
	Was the quality standard of the knowledge base determined?	Yes, the quality of the available knowledge was considered as reliable.
	Was an expert opinion assessment for suitability and appropriateness conducted?	No
Consistency	Were the basic scientific principles checked for consistency?	Yes, no inconsistencies were found.

	Are the knowledge base and the methodology used in line with the latest scientific knowledge and the state of the art?	Yes
	Were scientific limits determined?	Yes, a lack of understanding of intake, distribution, and depletion of ARS/SC in eel.
	Was it determined to what extent the scientific concepts and conclusions have already been reviewed in other fields of application?	No
	Was the empirical data used well-documented (internal and external validity, consistency of different sources)?	Yes, the German national consumption survey II (MRI, 2008) as well as the telephone interview on rarely consumed foods (Ehlscheid *et al.*, 2014) are both documented.
	How reliable (e.g. accurate, reproducible and stable over time) is the data used?	The data from the German national consumption survey II are from 2005-2007, while the telephone interview on rarely consumed foods was conducted in 2011. It cannot be excluded that consumer behavior has significantly changed until 2017.
Robustness	Can the data, assumptions and information be assumed to be reliable?	Yes, the German national consumption survey II as well as the telephone interview on rarely consumed foods are both scientifically respected.
	To which degree and in which direction do the identified data and knowledge gaps influence the result of the exposure assessment?	No exposure assessment was carried out.
	Can existing knowledge gaps have any major impact on the result?	Not applicable, since no result was obtained.
	Was the scientific knowledge base systematically reviewed and appraised in the context of the assessment problem?	No
	Has heterogeneity in published data and estimates been adequately recorded, diagnosed, documented and accounted for in the assessment procedures?	No
	What do we know about the transferability of exposure scenarios, models and data to the current application? How reliable will the results of such a transfer be?	Not applicable

	Is the degree of robustness of the scenario, the model used and the corresponding data high enough to ensure that a correct, plausible, and transparent result is obtained? Does this conform even for situations under unfavourable conditions (e.g. taking into account foreseeable misuse or possible errors in application)?	Not applicable

4.2.2.2. Question list for assessing the subjectivity of decisions

All questions in this list do not apply to this case study, since no exposure was estimated and therefore no assumptions regarding scenario, model, or parameter were made.

4.2.2.3. Question list for qualitative uncertainty analysis in relation to the definition of the task

Table 8: Question list relating to the definition of the task

Criteria	Questions[8]	Uncertainty assessment	Direction
Question formulation	Is the question formulation sufficiently precise for the purpose of exposure assessment?	Yes, consumer health assessment regarding marking of eel with ARS and SC.	0
Context	Is the application context of the exposure assessment described in sufficient detail?	Answer a Request from the German Federal Ministry of Food and Agriculture.	0
Protection perspective	Has it been defined at whose expense any residual uncertainty identified in the analysis is to be taken into account (consumer perspective, precautionary view, producer perspective/proof of risks)?	Yes, consumer health protection.	0
Population group to be protected	Is the population group to be protected defined with sufficient precision (e.g. individual persons, risk groups, special additional circumstances such as special consumption habits)?	Yes, population as a whole.	0
Protection goals	Is the subject of protection (e.g. irreversible health impairment, health impacts, change in taste, general purity criteria) defined clearly and described with sufficient precision?	Yes, avoidance of adverse effects on health are given as protection goals.	0
Protection level	Is the degree of the targeted protection levels (e.g. complete,	No, the degree of the targeted population is not specified.	--/++

	95% of the protection group, 95% of consumers or 95% of consumption events) defined clearly and with sufficient precision? What are specific sources and effects (on target variable) of uncertainties in the derivation of health based reference values?	 Not applicable, since no health based reference values have been derived so far.	
Scope and limitations	Are there uncertainties due to possible exclusion of questions, scenarios or parameters (e.g. nonconsideration of "background" exposure from the environment in the assessment of a specific product)? Do substitutes of the noxious agent exist that need to be taken into account?	None for ARS, but for SC other sources of exposure (in water, food, as well as in cosmetic products like toothpaste) are not taken into account. Not known	ARS: 0 SC: --

4.2.2.4. Question list for qualitative uncertainty analysis in relation to the exposure scenario

Table 9: Question list concerning the exposure scenario

Criteria	Questions	Uncertainty assessment	Direction
Development	Is the contaminant/agent that is the subject of assessment (hazard identification) defined with sufficient accuracy?	Yes: Alizarin Red S (ARS) (solfonated derivative of alizarin) and strontium chloride (SC) in form of strontium-chloride-hexahydrat.	0
	Do degradation products exist that need to be included in the exposure assessment?	No such information found.	-/+
	Does the noxious agent primarily occur in combination with other hazardous noxious agents so that it is to be viewed as the indicator substance of a group of noxious agents?	No such information found.	0
	Are the chemical, physical, biological and toxicological properties of the noxious agent adequately known?	No, the toxicological properties are mostly unknown.	---/+++
Release/	Are all primary sources of the	Yes	0

sources	noxious agent known? Is the complete material flow (e.g. quantity balance) of the noxious agent known in terms of development, spread and reduction? Are there multiple sources of the noxious agent that might occur in a correlated manner? Are processes of migration, release or crosscontamination possible?	 Typical values for the concentrations of ARS/SC in water are given in literature, but it cannot be excluded that in practice also other concentrations are used. For ARS, reduction (via binding with other ions in water) is reported, but degree of this effect not known). No Not applicable	 ARS:+ SC -/+ 0
Spread	Can the substance flow to the secondary contact media (air, drinking water, food, products) be fully traced and be described numerically? Are the exposure pathways (including the background contamination and the carryovers from other sources) fully taken into account? Are the routes of exposure that are to be taken into consideration clearly characterised? Can heterogeneous conditions of exposure be considered in summarised fashion through an aggregation of the influencing factors, by building groups of the products or food items, by generalisation of the represented life situation and by abstraction from the environmental conditions?	No, the amount of intake of ARS or SC for eel is unknown, as well as ARS/SC intake in edible tissue. In case of SC, the background contamination of water with strontium is not taken into account. Yes Not applicable	---/+++ Rated in Table 4 Scope and Restrictions 0
Reduction	Are the mechanisms by which a reduction of the concentration/amount of the noxious agent in the contact medium is possible (e.g. air exchange rate, mixing, degradation, decomposition) known and characterised?	No	?+

Contact: exposed population	Is the target population of the exposure assessment adequately described?	Yes, the whole population.	0
	Is the exposed population considered in the assessment identical with or sufficient similar to the target population described in the scope of the assessment? Which might be the major differences between the protection group and the definition of the target group of the exposure assessment?	No data for children younger than 14 years old. Children might have due to their low body weight a higher exposure per body weight.	---
	Are those groups or subgroups with special or excessive exposure behaviour taken into account adequately and described in detail?	No, sample size is too small for stratification. Especially, anglers are not explicitly taken into account.	--
Exposure events	Are the exposure events to be considered adequately described?	The measured exposure events do not discriminate whether European or other eel species are consumed. It is also unclear whether the eel was marked by ARS/SC or not.	++
	Is it possible to describe the exposure per exposure event?	Eel consumption is considered as per day, no differentiation in single exposure events.	0
Spatial, timebased and situational differences	Are the sources of exposure uniformly distributed for the exposed group (e.g. clearly defined technological processes of development, destruction or decontamination in the case of microorganisms)?	Eel consumption varies remarkably among the population.	0
	Are time-based and spatial differences (e.g. concentrations, intensities, short-term or seasonal changes, cycles, trends over time, climatic, regional or local differences, differences in lifestyles or modes of behaviour) and the microenvironment (e.g. pH level …) adequately defined?	Not applicable	
	Are the exposure conditions similar for the population under concern, for both genders and for all ages as	Not applicable	

	well as for difference life phases (e.g. school and work days, weekends and more specific pregnancy, hospitalisation …)?		
Risk management measures (RMMs)	Are the risk management measures (RMM) to be considered adequately described? Are RMMs adequately depicted for the scenarios used in the exposure assessment? Are all variables that might be influenced by known risk management measures (e.g. legal regulations, recommendations for application and usage) taken into account in the scenario description? Is noncompliant behaviour foreseen in the regulatory process (e.g. by communicated or noncommunicated RMMs[11]) for application and usage part of the exposure analysis?	Not applicable	

4.2.2.5. Question list for uncertainty analysis in relation to the choice of model

All questions in this list do not apply to this case study, since no model was created.

4.2.2.6. Question list for qualitative uncertainty analysis in relation to model parameters

Following quantities were treated as model parameters for uncertainty assessment, although no model was created:

1. ARS/SC concentration in edible eel tissue
2. Eel portion size for a consumption incidence
3. Frequency of eel consumption

In the following, they are referred to as 1., 2., and 3.

Table 10: Question list in relation to model parameters

Criteria	Questions	Uncertainty assessment	Direction
Expert opinions, default assumptions	Were default assumptions/expert opinions used for the value of parameters in the exposure estimates? If so, does the derivation of the default assumption/reference value (e.g. risk-covering or average, probable value) correspond to the	No	0

	objective and the step of the exposure assessment? Do agencies use deviating default values for the same parameter and if so, how can this be explained? Are the values plausible in terms of the objective?	 Not applicable	
Definition and quantification of the influencing variables	Does the model parameter meet the requirements of the exposure model (e.g. units of measurement, precision, stratifications, restrictions etc.) and adequately depict the range of variability for the subjects under examination?	Not applicable	
	Is the variable with its chosen value characteristics suitable for the description of the considered attributes of the target population?	1. Yes 2. Yes 3. Yes	0 0 0
	Do the characteristics of the time-based, spatial and inter-individual variations correspond to the exposure and risk model? What is the time interval (e.g. short-term, long-term, lifetime estimate, area under the curve, body burden indicators etc.) of the data?	Not applicable	
	Is the parameter of interest measured directly or calculated using conversion or assumptions of surrogate data? Are data available for the calibration and validation of the assumptions / conversion?	1. Not applicable 2. No, weight estimation via picture book. 3. Yes	 --/++ 0
	If only classified (interval-scaled or binned) data are available, is this classification sufficient for the purpose of the modelling? If parameters are derived from confidential data, is the level of information that can be provided sufficient to judge its adequacy?	1. Not applicable 2. Not applicable 3. Eel consumption frequency data is binned and considered suitable for quantification purpose.	 0
Reliability of meas-urements	Is the data collection method scientifically accepted and validated?	1. Not measured. 2. Yes, Two times 24h recall data from German national consumption survey II	---/+++ 0

		3. Yes, representative telephone interview with 1004 participants to collect information about rarely consumed foods during the previous year.	0
	Are the sources and the methods for data collection or measurement adequately documented in the literature?	1. Not applicable 2. Yes 3. Yes	 0 0
	Which bias and measurement errors might result from sampling and sample processing (e.g. contamination of the samples), analysis and the measuring methodology (e.g. calibration, quality assurance), determination and calculation of the model parameter (e.g. validation)?	1. Not applicable 2. 24h recall data is not much suitable for rarely consumed food (leads to small number of incidences). Moreover, it was not specified if European eel or other eel was consumed, which overestimates consumption of European eel. Furthermore, portion weight was estimated using a picture book.	 ---/+++
	Might the data, e.g. self-provided data from questionnaires, have systematic errors (over-/underreporting, bias due to social desirability)?	3. Questionnaires that ask consumers to report incidences of eel consumption over a long period of time may lead to underreporting, since incidences of consumption may have been forgotten. Moreover, it was not specified if European eel or other eel was consumed, which overestimates consumption of European eel.	--/++
	What are the possible consequences of the inclusion or exclusion of values below the detection or quantification limit? How were the values below the detection or determination limit quantified? How were missing values in the data set handled?	Not applicable	
	Were possible sources of systematic error and bias adequately discussed?	Mentioned above	
	Are there indications of widely differing values in the study? Do they point to special exposure conditions, missing influencing factors or "outliers"? Were "outliers" adequately handled? In the case of categorical data, is the diagnostic sensitivity and	1. Not applicable 2. No 3. No	 0 0

	specificity of the determination method or its positive/negative predictive value known and taken into account?		
Quality of data sources	Are data available from studies, systematic surveys or routine data?	1. No data 2. Study 3. Survey	 0 0
	Is the study protocol appropriate?	1. Not applicable 2. Yes 3. Yes	 0 0
	Was the study from which the data was taken performed with the aim of risk or exposure assessment?	1. Not applicable 2. No 3. Yes	 0 0
	Is the data set used original or secondary data?	1. Not applicable 2. Yes 3. Yes	 0 0
	Are there indications of different origins of the data in a study (e.g. different surveys, timeframes, laboratories, analysis methods etc.)? Was the resulting heterogeneity taken into account in the evaluation?	No such indications.	0
	Are there alternative studies on the same parameter that might confirm the quantification of the parameter choice(s)?	1. Not applicable 2. No 3. No	 0 0
	Is the study design adequately documented and in correspondence to pertinent scientific standards?	1. Not applicable 2. Yes 3. Yes	 0 0
	Is it likely that declared or undeclared interests unduly compromise the relevance or reliability of the data?	No	
Study population	Is the study population clearly defined?	2. Yes, general population (age 14-80) 3. Yes, general population (age 14 and older)	0 0

	Does the study cover all stratifications that are seen as important in order to take account of (for example) regional, climatic, time-based differences (e.g. seasonal variation, cycles, trends over time), different microenvironments (e.g. production, storage, packaging, preparation conditions), different lifestyles (e.g. activities, dietary requirements) etc.?	1. Not applicable 2. No, anglers are missing 3. No, anglers are missing	Rated in Table 5 Contact: exposed population
	Are there sufficient gender and age stratifications (e.g. babies, small children, children, adolescents, adults, seniors etc.)?	2./3. No children younger than 14 years included.	Rated in Table 5 Contact: exposed population
	Which selection effects may occur with a small sample sizes?	2. Uncertain: Over- or Underestimation 3. Uncertain: Over- or Underestimation	?-/+ ?-/+
	In which way would known biases associated with the respective study design affect the reliability of the data?	3. Underreporting of frequency of eel consumption and no diffentiation between eel species.	?-/+
Representativeness	Does the sampling strategy and the size of the sample ensure representativeness for the study population?	2. Yes, number of persons involved is large (13926) 3. Yes, number of people involved large (1004)	0 0
	Can results of the sample be transferred to the target population and the scope (regional, temporal) of the exposure assessment?	2./3. Results cannot be transferred to children younger than 14 years old.	Rated in Table 5 Contact: exposed population
		2. Study performed in 2005-2007 3. Survey performed in 2011	-/+ -/+
	Which assumptions and extrapolations are made, described and justified?	None	0
Details of correla-tions/dependencies	Have relevant correlations between influencing factors (e.g. consumption and body weight) been described and taken into account in the model (e.g. in-take/breathing rate/body surface	2. Yes individual consumption of eel per body weight 2./3. Correlation of frequency of eel consumption with amount of daily consumed eel not taken into account	0 ?-/+

	per kg body weight)? If there are correlations and structural dependencies, were they described in a transparent and logical way?	 Not applicable	
Evaluation methodology	**With deterministic estimates:** Are the statistical descriptions reported in a transparent and logical manner? Is the sample large enough to estimate the required parameters with sufficient precision? Which level of statistical precision (standard error of estimate, SEE and confidence intervals) has the exposure estimate? **With probabilistic estimates:** Are the statistical methods and selection criteria for distributions described in a transparent and logical way? Were considerations reported or additional data sets used to justify the selection of the distribution type? Is the sample size for the parameters considered large enough to accommodate the required distribution, especially extreme percentiles, with sufficient precision? Was the precision of the distribution fit and the corresponding parameters specified by providing confidence intervals, goodness-of-fit measures (e.g. Kolmogorov-Smirnov distance)? Were relevant statistical indicators (e.g. skewness, mean/median ratio, percentiles) of the empirical and the parametrically distribution compared	 1. Not applicable 2. Yes 3. Yes 2. Only 19 reported cases of eel consumption. 3. 1004 persons interviewed, but only 340 reported to have consumed eel during a one-year period. Therefore for heavy consumers the effective sample size is rather small. Not calculated Not applicable	 0 0 ---/+++ --/++

	and dis-cussed? Which assumptions were made to fit a distribution using small samples? What are the consequences of these assumptions for the target variable of the exposure assessment?		

4.2.2.7. Question list for qualitative uncertainty analysis in relation to the method of exposure calculation

All questions in this list do not apply to this case study, since no exposure was calculated.

4.2.3. Standardised qualitative presentation of the findings of uncertainty analysis for primary documentation

The following table summarizes the findings of the uncertainty analysis. As explained earlier, the subjectivity of choices could not be assessed. Moreover, the confidence of the knowledge base was only evaluated if appropriate. The symbols used where defined in Table 6 (see BfR guideline p.37 (Heinemeyer *et al.*, 2015)):

Table 11: Standardised presentation of the findings of uncertainty analysis

Identified aspects and magnitude of uncertainties in the exposure assessment			
	Degree of uncertainty	**Confidence in the knowledge base**	**Subjectivity of choices**
1. Goal and question formulation of the exposure assessment			
Question formulation	0		
Context	0		
Protection perspective	0		
Protected population	0		
Goals of protection	0		
Protection level	--/++		
Restriction of scope	ARS: 0 ;SC: --		
2. Exposure scenario			
Characterisation of the noxious agent	---/+++	Low	
Exposure source and origin, exposure routes and pathways (media)	ARS:+ SC: -/+	Medium	
Possible exposure paths	---/+++	Low	
Exposed groups of people/population	---	High	
Exposure events	++	Medium	
Assumed spatial, time-based and situational differences/lifestyles/modes of behaviour and microenvironment	0	Low	
Risk management measures	not applicable		
3. Exposure model			
Exposure estimator: definition	not applicable		

of the target variable				
Concept and assumptions used for the translation of the scenario into model equations	not applicable			
Dependencies/Correlations	not applicable			
Model structure, e.g. stratifications	not applicable			
Choice of model equation	not applicable			
Model extrapolation	not applicable			
Risk management measures	not applicable			
4. Parameters (to be completed separately for each parameter)				
	ARS/SC concentration in edible eel tissue	Daily amount of eel consumption in case of eel consumption	Frequency of eel consumption	
Expert opinions, default assumptions	0	0	0	
Definition, units and quantification of the influencing variables	not applicable	not applicable	not applicable	
Reliability of measurements	---/+++	---/+++	---/+++	Low|High|High
Quality of the data sources	not applicable	0	0	High
Study population	not applicable	?-/+	?-/+	High
Representativeness	not applicable	-/+	-/+	
Correlation structure	not applicable	?-/+	?-/+	Low
Evaluation methodology	not applicable	---/+++	--/++	
5. Documentation of the exposure calculation				
Deviations	not applicable			
Review of calculations	not applicable			
Deficient report compilation	not applicable			
Verification	not applicable			

4.2.4. Summary of the uncertainty analysis applying the BfR guidance

The uncertainty analysis has shown that large uncertainties are prevalent, which has prevented an exposure assessment in the original assessment. The dominant uncertainties are:

1. Missing knowledge about the toxicity of ARS and SC. Except the subchronic toxicity for SC, no scientifically sound knowledge is available.
2. Reliable data are missing for ARS/SC concentration in edible eel tissue. ARS/SC intake per eel and the fraction of ARS/SC which is actually accumulated in edible eel tissue are unknown. Moreover, it is uncertain whether ARS/SC is released from eel over time.
3. The data for eel consumption is not sufficient to allow for any stratification, since only 19 daily eel portion sizes were reported. This is a very small number even without any stratification. Furthermore, data for eel consumption was not available for children younger than 14 years old. But this group might belong to the high exposed consumers, since due to their smaller body weight, a larger exposure with regard to bodyweight may occur.

For the existing data sources of eel consumption, following important uncertainties could be identified:

- 24h recall data from the German National Dietary Survey II (MRI, 2008)

  - 24h recall data is not much suitable for rarely consumed foods due to the small number of incidences (19 in case of eel); this allows no stratifications (e.g. anglers or high consumers).
  - Due to the sample size of 19, large sample errors are to be expected.
  - Eel portion size is determined via comparison with a picture book, measurement errors are therefore to be expected.
  - No children under 14 years are included.
  - It is not differentiated whether European eel or other eel species are consumed; moreover the percentage of marked eel (either with ARS or SC) is unknown.
  - The existence and magnitude of correlation effects between frequency of eel consumption and eel portion size are not known.
- Telephone survey on rarely consumed foods:
  - No children under 14 years are included.
  - It is not differentiated whether European eel or other eel species are consumed; moreover the percentage of marked eel (either with ARS or SC) is unknown.
  - Since the telephone survey asked the participants to estimate their eel consumption frequency over a long time period (one year), effects of underreporting may occur.

It can be concluded that the BfR uncertainty analysis for the exposure assessment succeeded not only in identifying the main knowledge deficiencies, but also detected and assessed qualitatively the sources of uncertainty for the existing data. The uncertainty analysis gives helpful advice which additional data should be collected (e.g. data regarding ARS/SC intake in eel and fraction of ARS/SC accumulated in the edible tissue). Moreover, it also establishes a sound foundation for determining quantitative exposure (deterministic or probabilistic), since it outlines the knowledge gaps which needs to be accounted for (e.g. by model assumptions).

4.3. Uncertainty analysis applying the EFSA Guidance

4.3.1. Defining the Assessment Question

The German Federal Ministry of Food and Agriculture requested the BfR to assess the impact of the usage of Alizarin red S (ARS) or strontium chloride (SC) for marking of eel regarding consumer health protection. This question does not explicitly define certain points, which are listed below: Firstly, only exposure (in Germany) is assessed during this case study. Secondly, it was decided that other sources of SC exposure (e.g. in cosmetics) are not considered. Thirdly, the long term intake of (therefore relating to chronic toxicity) of ARS or SC will be estimated per year and per kg bodyweight.

4.3.2. Approach

The EFSA guidance document recommends a quantification of the overall uncertainty. For urgent assessments, EFSA advises to evaluate all uncertainties collectively for the assessment as a whole by using an expert knowledge elicitation (EKE). For this particular case study, we did not favour this option for the following reasons:

- A prior estimate of even a deterministic exposure value does not exist.

- The case study is complex due to the mostly unknown exposure pathway, like eel marker intake, accumulation of eel marker in edible eel tissue. This complexity makes it hard to assess an exposure estimate in one step for an overall EKE.
- Another effect this complexity is that the preparation of the EKE could require more time than available (e.g. for collecting the necessary information, assembling a variety of experts knowledgeable for different parts of the problem, thoroughly discussing different aspects of the exposure assessment).
- The existence of numerous and significant uncertainties with large magnitudes, like eel marker intake, accumulation of eel marker in edible eel tissue, eel consumption further complicates the matter.

Hence, we decided to carry out a quantitative assessment for each quantifiable source of uncertainty. This implies that an exposure model needs to be built first, which will already use a probabilistic description of some quantities. The resulting uncertainties are then combined by Monte Carlo simulations.

Given the short time given for the original assessment, the exposure of ARS/SC due to eel consumption was estimated for the most exposed group. If those obtained values for the intake are well below the critical values, the whole population can be assumed to be safe. Otherwise, a refinement (including surveys to obtain more data on eel consumption) would be necessary. Nevertheless, estimating the ARS/SC intake from eel consumption for the most exposed group seems for a first step to be a sensible plan. Due to limited data, the term "most exposed group" is translated into a scenario, which will be discussed at a later stage of this report.

It needs to be emphasised that for this uncertainty analysis literature was used that was not published at the time of the original risk assessment performed by the BfR.

4.3.3. Model

Roughly spoken, the model can be divided into the following parts:

1. ARS/SC intake per eel
    a. ARS/SC intake per eel during the eel marking.
    b. Fraction of ARS/SC that accumulates in edible eel tissue and not in the bones.
    c. ARS/SC release from the eel.
    d. Eel weight.
2. Eel consumption
    a. Frequency of eel consumption.
    b. Portion size.

The output Y of the model describes the intake of eel marker per kg bodyweight per year. Y can in principal be calculated as follows:

$$Y = I * M$$

where I denotes the intake of eel marker per consumed [kg] eel and M the yearly consumed European eel in [kg] per [kg] bodyweight.

4.3.3.1. ARS/SC intake per eel

Young eels (glass-eels) are placed into a solution containing either ARS or SC for a time period ranging from several hours up to one day. They are no available measurements indicating the amount of intake of eel marker per eel. Nevertheless, the following facts can be established:

1. The minimal possible intake is larger than 0.
2. The maximal possible intake is that the entire marker (ARS or SC) present in the solution is absorbed by the eels.

In the absence of further information, it is assumed that the distribution for ARS/SC intake follows a uniform distribution. This assumption is similar to the use of non-informative priors in Bayesian statistics, which reflects a lack of prior knowledge.

The cumbersome task remains to determine the maximal possible intake. Regarding the ARS/SC concentration, the following values seem to be widely used: 150 mg/l ARS ((Simon *et al.*, 2009), (Neukamm, 2009), (Caraguel *et al.*, 2015), (Kullmann *et al.*, 2017)) and 1g/l strontiumchlorid-hexahydrat ((Wickstrom & Sjoberg, 2014), (Neukamm & Hempel, 2017)), which translates to 328.6 mg/l strontium. In one source, different experimental designs regarding the amount of water and number of eels were tested with ARS (Kullmann *et al.*, 2017). The design that maximises the amount of water per eel was 10 m^3 water with 200 kg eel (8g per eel), which translates to 0.4 l per eel and therefore a maximal intake of ARS of 60 mg per eel. No indications have been found that for SC the study design would need to be different. If the same design is considered for SC, the maximal intake would be 131.4 mg strontium per eel.

Due to the design, the maximal possible intake of ARS/SC per eel scales with the available volume. It should be noted, though, that in reality, due to limitations in the water uptake of the eel, it would be expected that intake only depends on eel marker concentration and not on the available volume. This effect might lead to an overestimation of ARS/SC per eel.

The next step is the mathematical description of the part of the intake, which is accumulated in the edible eel tissue and not in the bones. Since the eel consumption refers only to the edible part of the eel, the ratio of ARS or SC content in edible eel tissue to ARS/SC content of the whole eel is chosen as parameter and denoted as λ. As outlined in the description of the case study, ARS forms complexes with calcium, and strontium is taken up similarly to calcium. Since most of the calcium in vertebrates is located in the bones, it is assumed that the maximal possible value of λ is 1 and the minimal one 0. Due to the lack of further knowledge, a uniform distribution between 0 and 1 is assumed as probability distribution for λ. A value of 1 for λ means that the edible eel tissue has an average ARS/SC content (regarding the whole eel). This is a conservative assumption and could be refined with more information.

Since no information is available for the release of ARS/SC from eel, it is assumed that no release occurs (conservative assumption).

Finally, to estimate the average weight of the eels at the time they are caught, a cautious approach is chosen. In most parts of Germany, the minimum length of eels for fishery is 50 cm. With the use of Fulton's condition factor K, defined as $K=\frac{weight[g]*100}{length^3[cm]}$, the eel weight can in principle be determined. In literature, different ranges for K have been reported: 0.17-0.22 in (Durif *et al.*, 2005), 0.16-0.18 in (Simon, 2007), and 0.19-0.21 in (Stein *et al.*, 2016). The smallest value K=0.16 g/cm^3 with a standard deviation of 0.026 g/cm^3 was reported for Jungfernsee (Simon, 2007) was chosen (leads to the largest ARS/SC intake). Since 50 eels were used, the mean value for K is known with following standard error: K=0.16±0.004 g/cm^3. A normal distribution reflecting the given mean and uncertainty was selected to describe K. Using an eel length of 50 cm, the mean weight G was determined to be G=200±5 g.

Let J denotes the total amount of eel marker (ARS or SC) per eel, J_1 the amount of eel marker in edible eel tissue per eel, G the total eel weight, and G_1 the weight of the edible eel tissue. λ can then be expressed as

$$\lambda = \frac{J_1/G_1}{J/G}.$$

Therefore, the product $J * \lambda$ can be expressed as

$$J * \lambda = \frac{J_1}{G_1} G .$$

The amount of eel marker consumed per [kg] eel, denoted as I, is then given by:

$$I = J * \frac{\lambda}{G}.$$

4.3.3.2. **Eel consumption**

The most exposed group is considered for the exposure assessment. But this choice is restricted to adolescents and adults, since children younger than 14 years are not included in the consumption studies. From Table 5, it is deducted that only a negligible part of the population eats eel more often than twice a week. Therefore, the scenario of consumption of eel twice per week is chosen. Regarding the 24h recall data, it needs to be stressed that only eel consumption was measured, but not specified whether European eel or any other species of eel. For the following calculations, it was assumed that all reported incidences of eel consumption refer solely to European eel and that all consumed eel is marked with ARS or SC respectively. Since only 19 incidences of eel consumption were reported in the National Survey II, only the average consumed daily eel portion weight (denoted as X) was taken into account. Taken together with the scenario of fixed eel consumption frequency mentioned above, this means that any kind of long-term variability regarding eel consumption is neglected. This approach further implies that there is no correlation between frequency of eel consumption and eel portion size.

Moreover, the amount of consumed eel is estimated by comparison with a picture book and is therefore subject to severe uncertainty. For the EPIC-SOFT picture book (which was used in the National Survey II), the typical relative standard deviation was found to be typically between 50-100% of the mean (Ambrus *et al.*, 2013). As a cautious assumption, the standard deviation was assumed to be 100%. Given that we look for a distribution (regarding eel portion weight), that allows only for positive arguments and is skewed to the right (the portion size cannot be underestimated by more than 100%, but overestimated by more than 100 %), the log-normal distribution was chosen. Moreover, the log-normal distribution is long-tailed, which allows for comparable large values of daily eel consumption and might therefore overestimate the probability for very large values of daily consumed eels. Once this uncertainty is added to the individual consumed eel per day (they are 19 incidences in total), it can be divided by the individual body weight. The individual body weight was determined with an accuracy of 0.1 kg. This typically translates to a relative uncertainty of 0.1-0.2 % and it was by expert judgement decided that this is negligible in comparison to the 100% relative uncertainty due to the estimation of portion size.

Since only 19 incidences of eel consumption were recorded, sampling errors are to be expected. To account for this, a non-parametric bootstrap with replacement was employed. Each bootstrap sample consists of 19 values randomly chosen from the original sample. Then, the mean is determined for each sample and as a result, a bootstrap distribution of daily eel consumption per bodyweight is obtained.

Multiplying the portion size per bodyweight X with the frequency of eel consumption (twice per week), the consumption of European eel per year per bodyweight (denoted as M) can be determined as:

$$M = X\frac{2 * 365.25}{7}.$$

4.3.4. List of sources of uncertainties and their description

In the following, all identified uncertainties are summarized in two tables together with a very short description. A more detailed description is given in the section above. First are the scenarios (uncertainties that could not been quantified) listed. These are the uncertainties where the risk assessor felt not able to quantify them. Due to the lack of knowledge of the underlying phenomena, this list is quite long. It needs to be stressed that the results obtained from this uncertainty analysis are only correct under the assumption that the following scenarios are valid.

Table 12: Sources of uncertainty - unquantified

Source of uncertainty – unquantified	Assumptions
ARS/SC concentration in solution	150mg/l; 1g/l (strontiumchlorid-hexahydrat); Typical values used in literature
Volume of marker solution per eel	400 ml; largest volume found in literature
Model design for intake of ARS/SC per eel	Model assumption that ARS/SC intake scales with available solution volume
Release of ARS/SC from the eel	No release (conservative assumption)
Frequency of eel consumption	For most exposed group: twice per week derived from Table 1
Correlation between frequency of eel consumption and amount of daily consumed eel in case of a consumption incidence	No correlation assumed
Variability of average daily consumed eel weight in case of eel consumption for individual consumers	Ignored;
Fraction of marked European eel on the overall eel consumption	100% (conservative assumption)
Eel length (proxy for eel weight)	50 cm; legal minimum length for fishery purposes in most parts of Germany
Fulton's condition factor K	K=0.16 g/cm^3 with standard deviation of 0.026 g/cm^3; smallest value found in literature

Following uncertainties were quantified:

Table 13: Sources of uncertainty - quantified

Source of uncertainty – quantified	Descriptions
ARS/SC intake per eel (given marker concentration and solution volume per eel)	Uniform distribution from 0 to maximal value (all marker is absorbed from the solution)
Ratio of ARS/SC content in edible eel tissue to ARS/SC content of the whole eel	Uniform distribution from 0 to 1
Reported consumed daily eel weight data (24 h Recall), derived by comparison with photographic atlas	Relative standard deviation of 100% assumed (upper boundary of (Ambrus *et al.*, 2013); Fit a log-normal distribution that reproduces mean (taken from observation) and variance

Sample of daily consumed eel weight data consists of only 19 values	Simple non-parametric bootstrap (with replacement) to yield a bootstrap distribution
Fulton's condition factor K	K=0.16 g/cm^3 with standard error of ±0.004 g/cm^3 using a normal distribution
Body weight	±0.1kg; deemed negligible by expert judgement

The overall uncertainty was calculated using Monte Carlo simulations using a large number of 10,000,000 runs.

4.3.5. Results of the quantitative uncertainty analysis

In the following, the probability density distribution of ARS (Figure 2) and SC (Figure 3) intake over one year are depicted. It should be noted though, that these results are only valid under the scenarios listed in the previous section and only valid for adults, since consumption data for children were not available.

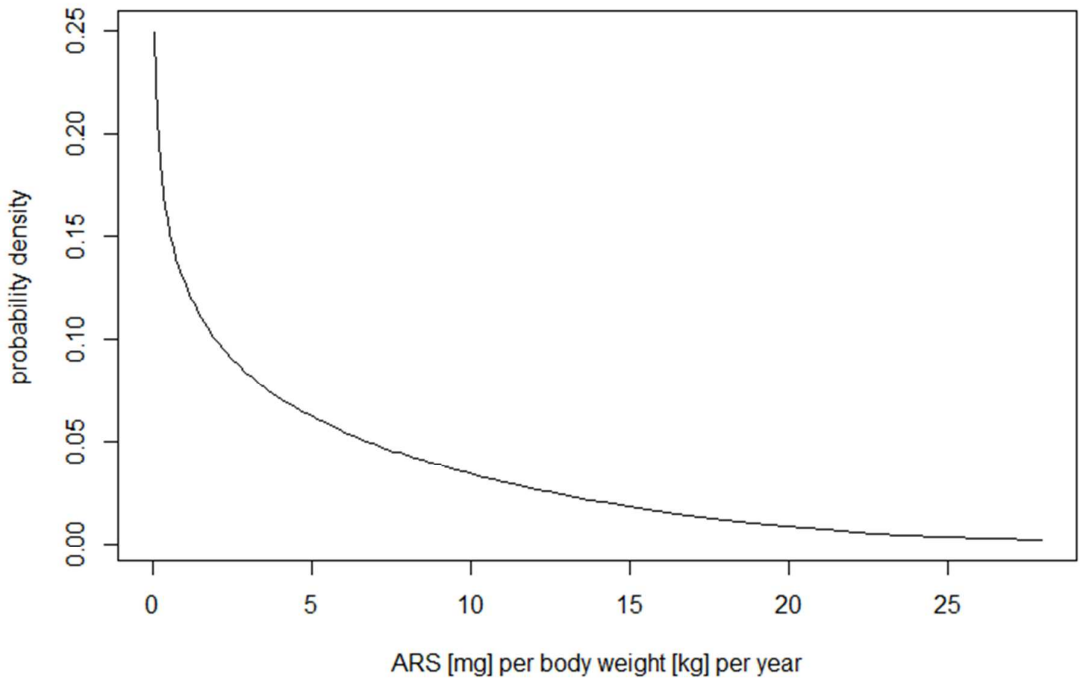

Figure 2: Probability density of ARS intake per kg bodyweight per year for a person consuming eel twice per week

The distribution for ARS intake has the following statistical features under the assumption of a person consuming eel twice per week (bw…bodyweight):

- 5% Percentil 0.2 mg/(kg bw year)

- Median: 4.7 mg/(kg bw year)
- Mean: 6.9 mg/(kg bw year)
- 95% Percentil 20.5 mg/(kg bw year)
- Standard deviation 6.9 mg/(kg bw year)

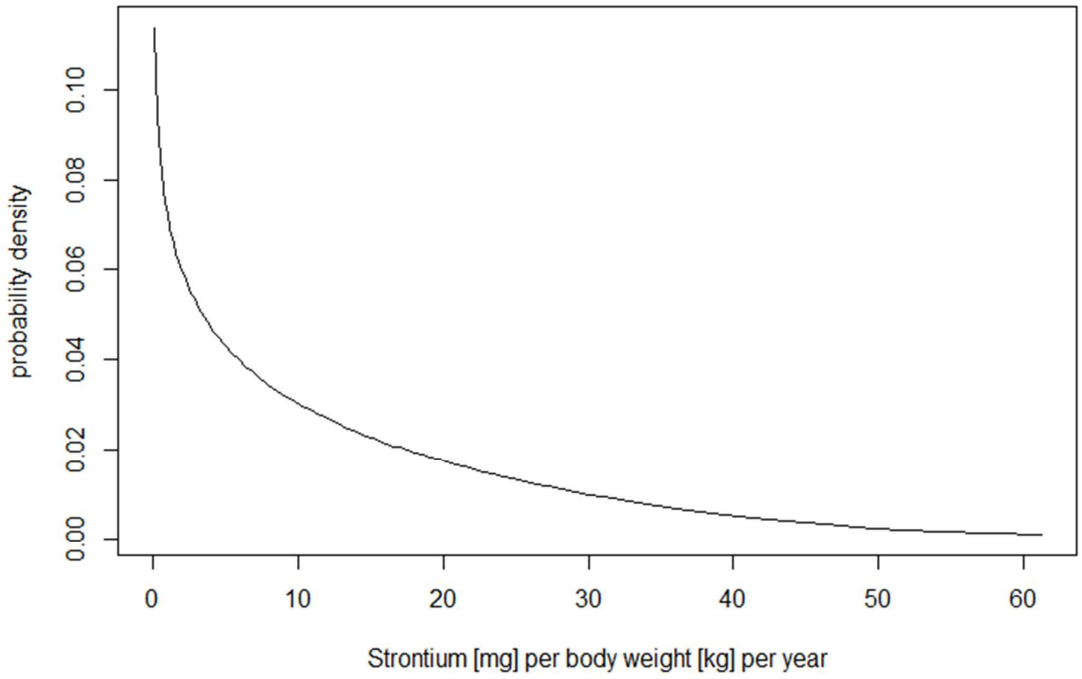

Figure 3: Probability density of strontium intake per kg bodyweight per year for a person consuming eel twice per week

The distribution for strontium intake has the following statistical features under the assumption of a person consuming eel twice per week (bw…bodyweight):

- 5% Percentil: 0.5 mg/(kg bw year)
- Median: 10 mg/(kg bw year)
- Mean: 15 mg/(kg bw year)
- 95% Percentil: 45 mg/(kg bw year)
- Standard deviation 15 mg/kg bw year)

It should be noted, that the probability density distribution diverges at 0 (for both ARS and SC), but the probability for zero intake is well defined and 0.

4.3.5.1. Sensitivity Analysis

In order to determine the individual contribution of the sources of uncertainty that are quantified within the model, a variance-based sensitivity analysis (also called Sobol method) was carried out (Saltelli *et al.*, 2010). This particular kind of sensitivity analysis constitutes a global sensitivity analysis, which allows a full exploration of the input space. The variance of the output of the model is decomposed into several parts, which can either be contributed to single input variables or to interactions between them. First of all, the first-order indices are determined. They indicate the relative contribution to the overall variance by only varying the particular input (e.g. source of uncertainty). Furthermore, the total total-effect indices are calculated, which are representing the relative reduction of the overall variance if the respective input remains constant. The estimation of both indices was carried out using Monte Carlo simulations (1,000,000 runs). The results are as follows:

Table 14: Sensitivity analysis concerning the expore of ARS and SC from eel

Source of uncertainty	First-order index	Total-effect index
ARS/SC intake per eel	0.33	0.49
Ratio of ARS/SC content in edible eel issue to ARS/SC content of the complete eel	0.33	0.49
Amount of consumed eel	0.13	0.24
Average eel weight of 50 cm long eel	0.0005	0.001

For the uncertainty of the daily consumed eel weight, two uncertainties needed to be combined: First, the uncertainty due to weight estimation via a picture book, and second, the sampling uncertainty (due to the sample size of 19 incidences). An individual treatment would have been computationally demanding, since the bootstrap of the 19 eel portion weights can only be carried out once the uncertainty due to inaccurate estimation of portion size is done. To get a representative bootstrap distribution, N simulation runs need to be carried out – for each individual realisation of weight estimation error. Since they are typically the same number N of runs required, in total N^2 simulation runs are necessary. Given that N was chosen to be 10^6, the problem is evident. The combined assessment of both uncertainties yields a yearly eel consumption of 0.046-0.174 kg per kg bodyweight (for the 2.5% and 97.5% respectively). Ignoring the measurement error of eel portion weight, the interval shrinks to 0.073-0.111, while ignoring sampling error leads to 0.056-0.146. It can be therefore concluded, that both uncertainties have a significant impact, but the influence of measurement error of eel portion weight is larger.

To summarize, the uncertainties regarding the ARS/SC intake per eel as well as the uncertainty regarding the accumulation of ARS/SC in edible eel issue dominate the other uncertainties quantified within the model. To reduce the uncertainty for this exposure assessment, investigations to obtain reliable knowledge about these factors take priority. It should be noted, that due to the choice of assumptions, no variability is present in Fig. 2 and Fig. 3. Moreover, it should be kept in mind that many uncertainties (Table 12) were not assessed qualitatively or quantitatively and therefore nothing can be stated about their contribution to overall uncertainty.

4.3.6. Summary of the uncertainty analysis applying the EFSA guidance

The obtained results and therefore the conclusions are dependent on the validity of the assumed scenarios:

- ARS concentration in water: 150mg/l; strontiumchloride-hexahydrat: 1g/l
- 400 ml solution per eel
- ARS/SC intake scales with available volume of the solution
- no excretion of ARS/SC from eel
- all consumed eel is European eel either marked with ARS (for the ARS assessment) or SC (for the SC assessment)
- eel consumption twice per week
- daily eel consumption (in case of consumption incidence) does not vary and can be replaced by its average
- no correlation between frequency of eel consumption and amount of daily consumed eel in case of a consumption incidence
- eel length 50 cm
- Fulton's condition factor K=0.16 ±0.026 g/cm^3 (standard deviation)

The following conclusions can be drawn:

1. The 95% percentile of yearly intake of ARS (strontium) is 20.5 mg/kg (45 mg/kg). The mean (±standard deviation) is about 6.9±6.9 mg/(kg year) (15±15 mg/(kg year))
2. However, the results might be also several orders of magnitudes smaller, given that the probability density distribution diverges at 0.
3. Investigation of intake and distribution of ARS/SC in eel takes priority to reduce the uncertainty of the considered uncertainties of the exposure assessment.
4. Obtaining further eel consumption data is necessary for stratification (e.g. to include children, to distinguish different consumption patterns,…).
5. Investigations regarding the sources of uncertainty of the unquantified uncertainties (Table 12) are necessary.

Since a sufficient toxicological assessment of ARS/SC is lacking, the obtained exposure results cannot be evaluated regarding health risk. But if a threshold value for the toxicity of ARS/SC could be established, two scenarios are possible:

1. The threshold value is much larger than the overwhelming part of the probability distribution of eel marker intake per bodyweight.

   In this case, the exposure assessment suffices and it can be concluded that regarding the intake of ARS/SC from the consumption of eel (alone) no health risk is to be expected.

2. A non-negligible part of the probability distribution is larger than the threshold value.

   In this case, the exposure assessment does not indicate whether a health risks exist or not due to the large uncertainties of ARS/SC intake. This large uncertainty is mainly caused by the uncertainties connected with ARS/SC intake per eel and the fraction of ARS/SC accumulated in the edible eel tissue, were uniform probability distributions have been assumed. For a

refinement of the uncertainty analysis, expert knowledge elicitations (EKEs) can be conducted in order to narrow down these probability distributions. EKEs enable to quantify expert knowledge of quantities, which would be hard to quantify otherwise. In Bayesian terms, such a refinement can be interpreted as the shift from non-informative priors to subjective priors.

4.4. Evaluation of both uncertainty analyses

The results of the uncertainty analysis have shown that the BfR guideline is well suited for identifying sources of uncertainties. This becomes especially clear for topics that are first and foremost not in the focus of a quantitative uncertainty analysis. One example is the protection level, which has not explicitly stated by the risk manager (as for this case study), it becomes the task of the risk assessor to choose one. The BfR question list regarding the definition of the task of the BfR guideline explicitly broaches this issue. On a broader level, both guideline documents are in some parts complementary, since the BfR uncertainty analysis delivers already a complete set of uncertainties (which are pre-assessed) that makes it easy to plan a quantitative uncertainty assessment (following EFSA's guidance document). It should be mentioned that for this case study, the uncertainty associated with the portion size determination (24h recall data of the German National Survey II) was initially not taken into account. A comparison with the BFR question list quickly revealed this deficiency. On the other hand, the need to quantify this uncertainty (for the EFSA guidance document) triggered a literature review that also helped in assessing the magnitude of this source of uncertainty for the BfR guideline.

By performing an uncertainty analysis following the respective BfR guideline, it was noticed that the same source of uncertainty will occur at different points of the question list, which can cause confusion for the evaluation of direction/magnitude of the source of uncertainty. A more clear division between the question lists concerning the exposure scenario and model parameters would be helpful.

Finally, the time needed to complete the uncertainty analysis using either the BfR guideline or the EFSA guidance document is compared. It should be noted that the time needed for documentation was not included, which nevertheless amounted to a very substantial part. Moreover, the person carrying out these uncertainty analyses did not have any prior practical experience with any of the guidelines.

Time required for uncertainty analysis using BfR guideline:	5 working days
Time required for uncertainty analysis using EFSA guidance document:	14 working days
Time required for information gathering:	6 working days
Time required for building and running the model:	4 working days
Time required for implementing and running the sensitivity analysis:	4 working days

As expected, a quantification of the overall uncertainty together with a sensitivity analysis takes significantly more time than the qualitative uncertainty analysis recommended by the BfR guideline. Of course, it needs to be emphasised that the results from the quantitative uncertainty analysis contain much more information than the qualitative one as shown in this case study.

It should be noted that the EFSA guidance document recommends for urgent assessments another approach in order to save time: to evaluate all uncertainties collectively for the assessment as a whole by using an expert knowledge elicitation (EKE). For this case study this option was not chosen, because the authors assumed that due to (1) complexity of the case study (2) together with several large sources of uncertainties would lead to very large ranges of possible exposure values if it would be quantified by using a single EKE, which might not give much practical benefits to the risk manager.

5. Second case study: Aluminium in cocoa and chocolate

5.1. Introduction to the second case study

5.1.1. Description of the case study

In 2017, the Federal Institute of Risk Assessment (BfR) conducted an exposure assessment for aluminium in cocoa and chocolate (BfR, 2017b). The exposure assessment was motivated by above average aluminium concentrations in these products. The reason for this contamination with aluminium is unclear. Up to now, there is no evidence that aluminium enters during the manufacturing process. Moreover, for freshly fermented cocoa beans, no aluminium could be detected. A hypothesis states that aluminium may enter while the cocoa beans are dried on banana leaves on the ground and get in contact with soil – which necessarily contains aluminium.

5.1.2. Summary of the initial BfR exposure assessment

The BfR exposure assessment concluded that for children (adults were considered too, but this group is not relevant here and therefore not mentioned) in the age group 0.5-5 years, a substantial part of the tolerable weekly intake (TWI) is already exhausted by the consumption of cocoa and chocolate. Assuming average aluminium concentrations yielded for heavy consumers (95th percentile) an intake of 34.7 % of the TWI. In another scenario, high aluminium concentrations (95th percentile) were assumed for one product group and average values for the remaining ones. Intake of aluminium yielded 49-57 % of the TWI for heavy consumers (95th percentile of consumers of the high-contaminated product group only). Given that other sources of aluminium intake exists (cosmetics, materials containing aluminium in contact with food, other foods containing aluminium), it was inferred that for a substantial part of the population, the TWI is exceeded in the long-term. It should be noted that adults were considered too, but the respective aluminium intake per bodyweight was notably smaller than for children. Thus, in the presented uncertainty analysis here, adults were not considered.

5.1.3. Used data

5.1.3.1. **Aluminum content data**

In total, 1646 single aluminium measurements in cocoa and chocolate were available, which were aggregated to seven product groups. 14 values were below the limit of quantification and set to 0. For chocolate-, nougat-, and cocoa-cream no direct measurements for aluminum content were carried out. Instead, the original exposure assessment used a point-estimate for the ratio of cocoa-powder in this product group (10%) and proceeded with aluminum content of cocoa-powder.

Table 15: Aluminium content in chocolate and cocoa products in mg/kg

	Sugar panned chocolate	Milk chocolate /baking chocolate	Chocolate icing/ chocolate sprinkles/ chocolate coating	Chocolate with fillings	Dark chocolate	Cocoa powder	Beverage powder containing cocoa powder
N	153	371	21	61	500	489	51
Survey period	2007-2014	2002-2014	2005-2014	2007-2014	2002-2015	2004-2014	2005-2015
Mean	23.6	25.5	33.2	22.7	49.8	152.1	33.6

Median	19.5	15.7	25.0	11.8	48.9	159.0	33.8
90th percentile	40.0	60.9	63.4	52.0	81.3	216.0	52.7
95th percentile	52.9	73.3	77.0	68.4	91.9	235.7	58.5
Maximum	110.0	135.0	80.0	204.2	162.3	544.0	91.4

5.1.3.2. Consumption data from VELS

The VELS survey ((Banasiak *et al.*, 2005), (Heseker *et al.*, 2003) investigated the dietary intake of small children (age 6 months until younger than 5 years, in total 816 persons) and was conducted in 2001/2002. Two times three days dietary diaries were available. The portion sizes were determined by weighing. For out-of-home consumption, the portion sizes were estimated. Small children that are still breast-fed were not included; therefore the sample size reduces to 732 persons.

5.2. Uncertainty analysis for the exposure assessment of aluminium in cocoa and chocolate

5.2.1. Approach

As result of the previous case study, it was found that the BfR guideline on uncertainty and the EFSA guidance on uncertainty complement each other very well. Therefore, the BfR document is used mainly to identify all sources of uncertainty present in the exposure assessment. The resulting list of uncertainties is then used to apply the EFSA guidance. Employing a simple sensitivity analysis, the uncertainties are prioritised and subsequently a plan how to assess the uncertainties devised. The uncertainty of the different parts is estimated and combined. Finally, not assessed uncertainties are considered collectively in an expert knowledge elicitation (EKE) to assess the overall uncertainty of the exposure assessment.

5.2.2. Assessment Question

To estimate the 2017 average long-term aluminium intake (chronic toxicity) by consumption of chocolate and cocoa products in 2017 for infants from age 0.5 years to less than 5 years (which are not breastfed) in Germany for the 95th percentile of the population specified above (in µg/(kg bw)/week). Further stratification of the described population is not desired.

The results can be compared with the tolerable weekly intake (TWI) of 1 mg/kg for aluminium (EFSA-AFC, 2008).

5.2.3. Details of the original exposure model

5.2.3.1. Inputs for the calculation model

1) Amounts of consumption for the following product groups for each participant from the VELS study
    - Sugar panned chocolate
    - Milk chocolate/ baking chocolate
    - Chocolate icing/ chocolate sprinkles/ chocolate coating
    - Chocolate with fillings
    - Dark chocolate

- Cocoa powder
- Beverages containing cocoa powder
- Chocolate-, nougat-, and cocoa-cream

2) Measured aluminium content in the consumed products in the same groups described above with exception of chocolate-, nougat-, and cocoa-cream, for which the aluminium content in cocoa powder was used

3) Fraction of cocoa powder in chocolate-, nougat-, and cocoa-cream

4) Body weight of the participants

5.2.3.2. **Original exposure model**

In order to derive the intake of aluminium by consumption of cocoa and chocolate, following formula was used:

$$I = \frac{Consumption(per\ week) * Aluminium\ content}{bodyweight}$$

One scenario considered by the initial BfR exposure assessment was that consumers eat low-contaminated as well as high-contaminated products (regarding aluminium) such that in the long-term, average aluminium content can be assumed. Average, Median, and the 95th percentile of the complete population of the VELS study (except breastfed children) as well as only consumers of certain product groups were determined. Considered was also the case that consumers may eat products containing high values of aluminium. For this, for one product group the 95th percentile of aluminium content was taken, for the other product groups mean aluminium content assumed and only consumers of the product group with large aluminium content were considered. Mean, Median, and 95th percentile of the aluminium intake for such a group were estimated.

5.2.4. Question lists of the BfR guideline for identifying sources of uncertainty

The BfR guideline will be applied to the original exposure assessment under the scope of the new assessment question. The three dimensions of uncertainty are considered:

1) Degree of uncertainty
2) Appraisal of the knowledgebase
3) Subjectivity of choices

The appraisal of the knowledgebase as well as the subjectivity choices is for reasons of clarity placed first in the Tables 17 and 18. They are applicable to all steps of the exposure pathway, and their assessment (rating in low, medium, and high) is given in Table 24. The remaining question lists deal with the degree of uncertainty for the definition of the task (Table 19 in section 5.2.4.3), the exposure scenario (Table 20 in section 5.2.4.4), the model choices (Table 21 in section 5.2.4.5), the model parameters (Table 22 in section 5.2.4.6) and the procedure of exposure calculation (Table 23 in section 5.2.4.7). The degree of uncertainty is assessed for both, magnitude and direction of the uncertainty. The following symbols are used for assessing the uncertainties defined in Table 16 (BfR guideline p.37 (Heinemeyer *et al.*, 2015).

Table 16: Ordinal scale for classification of uncertainty

Degree of potential effect	Underestimation	Direction not known	Overestimation
Not discernible/ Negligible	0: Uncertainty has no discernible or a negligible effect on estimation of the risk	0: Uncertainty has no discernible or a negligible effect on estimation of the risk	0: Uncertainty has no discernible or a negligible effect on estimation of the risk
Low	-: Uncertainty can result in a low underestimation of the risk	-/+: Uncertainty can result in a low deviation in the estimation of the risk in both directions	+: Uncertainty can result in a low overestimation of the risk
Moderate	--: Uncertainty can result in a moderate underestimation of the risk	--/++: Uncertainty can result in a moderate deviation in the estimation of the risk in both directions	++: Uncertainty can result in a moderate overestimation of the risk
High	---: Uncertainty can result in a high underestimation of the risk	---/+++: Uncertainty can result in a high deviation in the estimation of the risk in both directions	+++: Uncertainty can result in a high overestimation of the risk
Not known	?-: Uncertainty can result in a underestimation of the risk of unknown magnitude	?-/+: Uncertainty can result in a deviation in the estimation of the risk in both directions and of unknown magnitude	?+: Uncertainty can result in a underestimation of the risk of unknown magnitude

A summarized overview of the degree of uncertainty will be given in Table 24 in section 5.2.4.8.

5.2.4.1. Question list for assessing the knowledge base

Table 17: Question list for assessing the knowledge base

Criteria	Questions	Assessment
Completeness	Was the essential and relevant knowledge base compiled in a manner that is necessary to obtain an exposure assessment with the desired accuracy?	Yes, the knowledge base was compiled in a manner that allows obtaining an exposure estimate.
	Were the most important deficiencies in the knowledge base identified?	The most important deficiency is that it is not known why aluminium in chocolate and cocoa in these concentrations occur and from which sources it enters. Furthermore, there is a lack of information about changes of consumption habits since 2001/2002 to 2017
	Were the possible effects of these weak points on the result of the exposure assessment controlled? Were assumptions identified that can compensate the weak points of the knowledge	Different scenarios (average aluminium concentration vs. brand-dependence of aluminium concentration together with brand loyalty) were employed

	base? Were all parameter values and results controlled using comparative calculations? Were all dependencies and interrelationships between model variables reviewed?	 No, only results of aluminium intake were compared with values obtained in France. Dependencies in food consumption were taken into account.
Reliability	Was the knowledge base checked for factual and methodological justifications? Was the knowledge base reviewed to ensure it is scientifically up-to-datedness? Was the quality standard of the knowledge base determined? Was an expert opinion assessment for suitability and appropriateness conducted?	The VELS study as well as the aluminium content measurements is a scientifically acknowledged method. The data from the VELS study are from 2001-2002. Alumnium measurements are from 2002-2015. Yes, the quality standard of the knowledge base was deemed sufficient. No
Consistency	Were the basic scientific principles checked for consistency? Are the knowledge base and the methodology used in line with the latest scientific knowledge and the state of the art? Were scientific limits determined? Was it determined to what extent the scientific concepts and conclusions have already been reviewed in other fields of application? Was the empirical data used well-documented (internal and external validity, consistency of different sources)? How reliable (e.g. accurate, reproducible and stable over time) is the data used?	Yes, the basic scientific principles are consistent. Yes the knowledgebase and the methodology used are in line with the latest scientific knowledge. No No Yes, the data from the VELS study is well documented as well as aluminium content data. The VELS data is from 2001-2002.
Robustness	Can the data, assumptions and information be assumed to be reliable? To which degree and in which direction do the identified data and knowledge gaps influence the result of the exposure assessment? Can existing knowledge gaps have any major impact on the result? Was the scientific knowledge base systematically reviewed and appraised in the context of the assessment problem?	Yes, the VELS consumption study can be assumed to be reliable. If aluminium content is brand-dependent, the 95th percentile of aluminium will be higher (given the existence of brand-loyalty). Use of scenarios (brand dependence + brand-loyalty vs. mean concentrations) to account for bandwidth of possibilities. Yes, the scientific knowledge base was reviewed in the context of the assessment problem.

	Has heterogeneity in published data and estimates been adequately recorded, diagnosed, documented and accounted for in the assessment procedures?	No heterogeneity identified so far.
	What do we know about the transferability of exposure scenarios, models and data to the current application? How reliable will the results of such a transfer be?	It is unclear if the results of the consumption study VELS (data from 2001-2002) are still valid in 2017.
	Is the degree of robustness of the scenario, the model used and the corresponding data high enough to ensure that a correct, plausible, and transparent result is obtained? Does this conform even for situations under unfavourable conditions (e.g. taking into account foreseeable misuse or possible errors in application)?	Yes, the robustness of the scenario, the model used and the corresponding data are robust enough to ensure a correct, plausible, and transparent result is obtained.

5.2.4.2. Question list for assessing the subjectivity of choices

Table 18: Question list for assessing the subjectivity of choices

Criteria	Questions[8]	
Scope of possible alternatives (decision-scope)	Were all possible alternatives for the selection scenarios, models or parameter described?	The considered age group (0.5-5 years), the decision to use the 95th percentile of population and to limit the assessment to cocoa and chocolate products were done by the risk assessor.
Differences between decisions of experts and stakeholders	Were concurrence and differences between the positions of different experts and/or stakeholders described?	Not applicable
Influence of situation-based restrictions on the decision	Was the influence of limited resources (e.g. research funds, infrastructure, working time for analysis and document preparation) on the selection decision determined?	The VELS data is from 2001/2002 and therefore not up to date, but the data of a new dietary survey (KiESEL) was not available yet.
Choice is guided by interests and values of the expert or stakeholder	Were possible effects of interests or scientific positions assessed with respect to procedural decisions? Is it to be assumed that the procedure may be guided by interests (e.g. for the application of specific methods and technologies)?	No conflicting interests.
Influence of the decision on the result of the exposure assessment	Was the influence of the choice of scenarios, models, and specific parameters on the result of the exposure assessment determined in a comparative manner?	Yes.

5.2.4.3. Question list for qualitative uncertainty analysis in relation to the definition of the task

Table 19: Question list relating to the definition of the task

Criteria	Questions[8]	Uncertainty assessment	Direction
Question formulation	Is the question formulation sufficiently precise for the purpose of exposure assessment?	Consumer health assessment regarding aluminium intake via chocolate and cocoa. Further specifications done by risk assessors.	? -/+
Context	Is the application context of the exposure assessment described in sufficient detail?	It is not known how aluminum enters into cocoa and chocolate	? -/+
Protection perspective	Has it been defined at whose expense any residual uncertainty identified in the analysis is to be taken into account (consumer perspective, precautionary view, producer perspective/proof of risks)?	Consumer health protection.	0
Population group to be protected	Is the population group to be protected defined with sufficient precision (e.g. individual persons, risk groups, special additional circumstances such as special consumption habits)?	The population group to be protected (children age 0.5-5 years) was chosen by the risk assessor. Assumption that infants are due to their small body weight more sensitive to the intake of toxic substances.	-/+
Protection goals	Is the subject of protection (e.g. irreversible health impairment, health impacts, change in taste, general purity criteria) defined clearly and described with sufficient precision?	Yes, adverse long-term effects on health (chronic) are to be avoided.	0
Protection level	Is the degree of the targeted protection levels (e.g. complete, 95% of the protection group, 95% of consumers or 95% of consumption events) defined clearly and with sufficient precision? What are specific sources and effects (on target variable) of uncertainties in the derivation of health based reference values?	No, the degree of the targeted population is not specified. The choice of the 95th percentile of the protection group (children age 0.5-5 years – whole population) was made by the risk assessor. Within this scope, all infants were taken into account, not only consumers, since over long-term nearly everyone consumes cocoa/chocolate products. Not applicable	--/++
Scope and limitations	Are there uncertainties due to possible exclusion of questions, scenarios or parameters (e.g. nonconsideration of "background" exposure from the environment in the assessment of a specific product)? Do substitutes of the noxious agent	Other sources of aluminium intake (e.g. cosmetics, food in contact with baking sheets made from aluminium) are not taken into account. Not applicable	---

	exist that need to be taken into account?		

5.2.4.4. Question list for qualitative uncertainty analysis in relation to the exposure scenario

Table 20: Question list concerning the exposure scenario

Criteria	Questions	Uncertainty assessment	Direction
Development	Is the contaminant/agent that is the subject of assessment (hazard identification) defined with sufficient accuracy?	Yes, aluminium and compounds containing aluminium.	0
	Do degradation products exist that need to be included in the exposure assessment?	No	0
	Does the noxious agent primarily occur in combination with other hazardous noxious agents so that it is to be viewed as the indicator substance of a group of noxious agents?	No	0
	Are the chemical, physical, biological and toxicological properties of the noxious agent adequately known?	For the exposure assessment, the knowledge of the properties is sufficient.	0
Release/ sources	Are all primary sources of the noxious agent known?	The primary sources of aluminium in chocolate and cocoa are not known for sure – it could be from the soil as well as from the manufacturing process.	? -/+
	Is the complete material flow (e.g. quantity balance) of the noxious agent known in terms of development, spread and reduction?	The material flow of aluminium is not known.	---/+++
	Are there multiple sources of the noxious agent that might occur in a correlated manner?	Since the primary sources of aluminium (that enters cocoa and chocolate) are not known, this question cannot be answered.	? -/+
	Are processes of migration, release or crosscontamination possible?	Possible, but not known.	? -/+
Spread	Can be the substance flow to the secondary contact media (air, drinking water, food, products) be fully traced and be described numerically?	No, explained in previous section	See section above
	Are the exposure pathways	Regarding the defined scope, the	0

	(including the background contamination and the carryovers from other sources) fully taken into account? Are the routes of exposure that are to be taken into consideration clearly characterised?	exposure pathway is given by the consumption of cocoa and chocolate.	
	Can heterogeneous conditions of exposure be considered in summarised fashion through an aggregation of the influencing factors, by building groups of the products or food items, by generalisation of the represented life situation and by abstraction from the environmental conditions?	For aluminium content, similar products were aggregated to product groups. However, since the contamination route is unclear, whether this is valid remains unknown.	? -/+
Reduction	Are the mechanisms by which a reduction of the concentration/amount of the noxious agent in the contact medium is possible (e.g. air exchange rate, mixing, degradation, decomposition) known and characterised?	Not applicable, since no reduction	
Contact: exposed population	Is the target population of the exposure assessment adequately described?	Yes, the 95th percentile of all children between 0.5 and less than 5 years which are not breastfed in Germany.	0
	Is the exposed population considered in the assessment identical with or sufficient similar to the target population described in the scope of the assessment??	Yes	0
	Which might be the mayor differences between the protection group and the definition of the target group of the exposure assessment?	Not applicable	
	Are those groups or subgroups with special or excessive exposure behaviour taken into account adequately and described in detail?	No further stratification for children between 0.5 and less than 5 years. Behaviour of children above the 95th percentile is not taken into account.	Not in the scope of assessment
Exposure events	Are the exposure events to be considered adequately described?	Consumption of cake and cereals containing chocolate or cocoa as well as ready-to-drink milk-mixed beverages containing cocoa are not considered	--
	Is it possible to describe the exposure per exposure event?	Consumption of chocolate and cocoa is considered as per day, no differentiation in single exposure events.	0

Spatial, timebased and situational differences	Are the sources of exposure uniformly distributed for the exposed group (e.g. clearly defined technological processes of development, destruction or decontamination in the case of microorganisms)?	Consumption of chocolate and cocoa varies remarkably among the population.	0
	Are time-based and spatial differences (e.g. concentrations, intensities, short-term or seasonal changes, cycles, trends over time, climatic, regional or local differences, differences in lifestyles or modes of behaviour) and the microenvironment (e.g. pH level …) adequately defined?	Seasonality of consumption of chocolate was not accounted for. Consumption in winter is likely higher than in summer.	--/++
		Possible changes in consumption of chocolate and cocoa from 2001/2002 to 2017 were not accounted for.	--/++
	Are the exposure conditions similar for the population under concern, for both genders and for all ages as well as for difference life phases (e.g. school and work days, weekends and more specific pregnancy, hospitalisation …)?	Dietary habits of children in the age group of 0.5 to 5 years vary remarkably.	Out of scope of assessment
Risk management measures (RMMs)	Are the risk management measures (RMM) to be considered adequately described? Are RMMs adequately depicted for the scenarios used in the exposure assessment? Are all variables that might be influenced by known risk management measures (e.g. legal regulations, recommendations for application and usage) taken into account in the scenario description? Is noncompliant behaviour foreseen in the regulatory process (e.g. by communicated or noncommunicated RMMs) for application and usage part of the exposure analysis?	Not applicable	

5.2.4.5. Question list for uncertainty analysis in relation to the choice of model

Table 21: Question list in relation to the choice of model

Criteria	Questions	Uncertainty assessment	Direction
Estimation of exposure : definition of the target variable	Are the variables of the exposure modelling process described with sufficient accuracy (e.g. mean/cumulative/maximum dose/concentration, unit, external/internal exposure, exposure events etc.)?	Number of aluminum measurements for some product groups not enough to estimate e.g. the 99th percentile, parametric distribution necessary.	-/+

	Does the exposure assessment (e.g. with respect to the units of the target variable, comparability of the calculation, reproducibility etc.) meet the requirements for (quantitative) risk characterisation[13] (e.g. TDI, ARfD)?	Yes, the result from the exposure assessment allows the comparison with the tolerable weekly intake.	0
	Does the calculation of exposure might confirm the achievement of the protection goals (e.g. compliance with exposure limits for children) for time-based or spatial frameworks?	No	
	Are there any alternative concepts for exposure estimates (e.g. duplicate diet studies, human biomonitoring)?	Given time and resources - no.	
Concept and assumptions for transfer of the scenario into mathe-matical model	Does the model equation deliver averages or extreme estimates as described in the scenario (scope of interest)?	The model can capture the averages as well as extremes.	0
	Was the aim of the choice of model the deliberate overestimation of the target value, and, if so, how great is the resulting overestimation?	No	0
	If yes, what are advantages and disadvantages in terms of uncertainties resulting from the use of distributions for the model parameters?	Not applicable	
Connections/ Correlations	Are there correlations and structural dependencies between the influencing variables described? Are they accounted for in the model?	Food intake is typically correlated, but taken into account in the model by default (use of comprehensive dietary study).	0
	In the event of multiple sources of the same noxious agent, for example, are there sources that occur in combination or correlation? To what extent and in what direction would the effect of non-consideration of correlations and interdependence affect the result?	Out of scope of assessment question.	
Model structure, e.g. stratifications	Are sufficient stratifications present in the model to take account of regional (e.g. climatic, region type, change of location, trade flow), time-based differences (e.g. seasonal, cycles, trends), different microenvironments (e.g. production, storage, packaging, preparation conditions), different lifestyles (e.g.	Seasonal consumption of chocolate was not taken into account.	Rated under scenarios

	activities) etc.?		
	Does the model contain sufficient gender and age stratifications (e.g. neonates, toddlers, children, adolescents, adults, seniors etc.)?	No further stratifications for infants (age 0.5 years to < 5 years) employed.	Out of scope for assessment question
	Are particularly exposed persons (e.g. specific dietary needs or following incorrect use of a product) taken into account by the model?	No particularly exposed group identified.	0
	Are the requirements for all model parameters of modelling described with sufficient precision (e.g. unit, precision, stratifications, restrictions etc.)?	Yes	0
Choice of model equation	Is the application of the model accepted by experts, tested or validated?	Standard model equation for intake estimations	0
	Does the model contain all the influencing factors of the exposure scenario?	Yes	0
	Is the applied formula generally scientifically accepted?	Yes	0
	Are all components and influencing factors of the model substantiated and explained?	Yes, two scenarios are employed for investigating the possibility that aluminium content is and is not brand dependent.	0
	Are assumptions transparent and described in terms of their influence on the target variable?	Assumptions regarding parameters exist and are mentioned there.	
	What is the quality (e.g. goodness of fit, considered influencing factors, restrictions) of model? Were the statistical methods for evaluation adequately substantiated?	The model is adequate.	0
	Does the degree of detail of the model correspond to that of the scenario?	Yes	0
	Does the model adequately consider the relevant processes in the exposure pathway (e.g. transformations, growth, degradation processes)?	Not applicable	
	Does the model correctly depict the relationships between all influencing variables (e.g. age, behaviour) and exposure factors (e.g. consumption frequency, water intake per body	Yes, intake per individual body weight is estimated.	0

	weight) that are seen as being relevant?		
	Are there conversions or decision variables (e.g. restricting parameter range limits) in the model that are disputable?	Not applicable	
	Were all paths and exposure sources taken into account in the model formulas?	Some foods were not taken into account, like cakes/cereals containing chocolate, ready-to-drink mixed milk beverages containing cocoa	(--/++) Rated under scenario
	Do the model equations adequately reflect the exposure process, in particular individual exposure events with regard to time- and path-specific correlations (e.g. habits of food consumption on the day of the week, season or festive days)?	Seasonal consumption of chocolate was not taken into account.	(--/++) Rated under scenario
	Is the model complexity balanced between the number of necessary influencing factors and the usage of assumptions for information gaps? Which assumptions are made?	Assumption that cocoa content of cocolate cream is 10%	-/+
	Are there alternative model proposals published?	No	0
Extrapolations of the model	Was the model adopted as in analogy from another application?	No	0
	Does the application of the model extrapolate to new areas for the scenario?	Long-term intake was estimated by the 6-day dietary study of VELS. Overestimates variance and therefore the 95th percentile of consumption.	++
	Is the model used with parameters for which it initially was not designed or evaluated; (e.g. for the evaluation of time trends or for a different degree of local aggregation)?	No	0
Risk management measures	Are all variables that can be influenced by risk management measures (e.g. legal regulations) taken into account in the model? Are parameters describing such mitigation options selected in accordance with observed, expected or intended practice?	Not applicable	

5.2.4.6. Question list for qualitative uncertainty analysis in relation to model parameters

In order to give a comprehensive overview, all parameters are considered in one question list. The following parameters are considered:

1) Amount of foods consumed (containing cocoa/chocolate)
2) Aluminium content of these foods
3) Cocoa content of chocolate cream
4) Body weight

Table 22: Question list in relation to model parameters

Criteria	Questions	Uncertainty assessment	Direction
Expert opinions, default assumptions	Were default assumptions/expert opinions used for the value of parameters in the exposure estimates? If so, does the derivation of the default assumption/reference value (e.g. risk-covering or average, probable value) correspond to the objective and the step of the exposure assessment?	No	0
	Do agencies use deviating default values for the same parameter and if so, how can this be explained? Are the values plausible in terms of the objective?	Not applicable	
Definition and quantification of the influencing variables	Does the model parameter meet the requirements of the exposure model (e.g. units of measurement, precision, stratifications, restrictions etc.) and adequately depict the range of variability for the subjects under examination?	1.-4. Yes	0
	Is the variable with its chosen value characteristics suitable for the description of the considered attributes of the target population?	1.-4. Yes	0
	Do the characteristics of the time-based, spatial and inter-individual variations correspond to the exposure and risk model? What is the time interval (e.g. short-term, long-term, lifetime estimate, area under the curve, body burden indicators etc.) of the data?	1. VELS survey consisted of 6 days, but long-term intake is of interest. The simple extrapolation between both time periods overestimates variability and overestimates therefore aluminum intake of the 95th percentile.	++
	Is the parameter of interest measured directly or calculated using conversion or assumptions of surrogate data? Are data available	1.-4. Aluminium content was measured directly (except chocolate cream) as well as amount of foods consumed and body weight. Cocao content of chocolate	0

	for the calibration and validation of the assumptions / conversion? If only classified (interval-scaled or binned) data are available, is this classification sufficient for the purpose of the modelling? If parameters are derived from confidential data, is the level of information that can be provided sufficient to judge its adequacy?	cream from product description 2. Aluminium content of chocolate cream was estimated by using aluminium content of cocoa powder and cocoa content of chocolate cream Not applicable	0
Reliability of meas-urements	Is the data collection method scientifically accepted and validated?	1.-4. Yes, the VELS study, the aluminium measurement method, and weighing of body weight are scientifically accepted.	0
	Are the sources and the methods for data collection or measurement adequately documented in the literature?	1.-4. Yes	0
	Which bias and measurement errors might result from sampling and sample processing (e.g. contamination of the samples), analysis and the measuring methodology (e.g. calibration, quality assurance), determination and calculation of the model parameter (e.g. validation)? Might the data, e.g. self-provided data from questionnaires, have systematic errors (over-/underreporting, bias due to social desirability)?	1. The VELS study contains two three-days diaries of dietary intake. Since the long-term intake was calculated, intrapersonal variability in dietary habits as well as the seasonality of chocolate consumption will lead to an overestimation for high consumers.	(++) Rated under scenario and above
		1. Foods like muffins, chocolate cake, ready-to-drink milk-mixed beverages containing cocoa were not included in the original exposure assessment.	(--) Rated under scenario
		1. Most foods are weighted (accuracy 1 g), but out of home consumption was only estimated.	-/+
		2. The single products used for the aluminium content measurements are not necessarily representative regarding consumption.	-/+
		2. Measurement errors occur for aluminium content measurements.	-/+
		2. 14 from 1646 samples of the aluminium measurements were below the limit of quantification. These values were set 0. No missing values.	0
		3. For chocolate cream, no aluminium content was available, therefore the cocoa powder content was estimated and its aluminium content taken for further estimation.	-/+

		4. The body weight was estimated with a precision of 200g.	0
	What are the possible consequences of the inclusion or exclusion of values below the detection or quantification limit? How were the values below the detection or determination limit quantified? How were missing values in the data set handled?	Mentioned above	
	Were possible sources of systematic error and bias adequately discussed?	Mentioned above	
	Are there indications of widely differing values in the study? Do they point to special exposure conditions, missing influencing factors or "outliers"? Were "outliers" adequately handled? In the case of categorical data, is the diagnostic sensitivity and specificity of the determination method or its positive/negative predictive value known and taken into account?	No such indications found.	0
Quality of data sources	Are data available from studies, systematic surveys or routine data?	1./4. Dietary intake of infants (0.5 to <5years) comes from the VELS study as well as body weight determination. 2. Measurements of aluminium content were regularly measurements of the Federal Office of Consumer Protection and Food Safety.	
	Is the study protocol appropriate?	1./4. Yes	0
	Was the study from which the data was taken performed with the aim of risk or exposure assessment?	1./4. Exposure assessment.	
	Is the data set used original or secondary data?	1.-4. Original data.	
	Are there indications of different origins of the data in a study (e.g. different surveys, timeframes, laboratories, analysis methods etc.)? Was the resulting heterogeneity taken into account in the evaluation?	2. Aluminium measurements conducted by different laboratories.	0
	Are there alternative studies on the same parameter that might confirm the quantification of the parameter choice(s)?	No	

	Is the study design adequately documented and in correspondence to pertinent scientific standards?	Yes	0
	Is it likely that declared or undeclared interests unduly compromise the relevance or reliability of the data?	No	0
Study population	Is the study population clearly defined?	1./4. Yes, infants 0.5 to <5 years old which are not breastfed.	0
	Does the study cover all stratifications that are seen as important in order to take account of (for example) regional, climatic, time-based differences (e.g. seasonal variation, cycles, trends over time), different microenvironments (e.g. production, storage, packaging, preparation conditions), different lifestyles (e.g. activities, dietary requirements) etc.?	1./4. Yes	0
	Are there sufficient gender and age stratifications (e.g. babies, small children, children, adolescents, adults, seniors etc.)?	Not relevant for assessment question.	
	Which selection effects may occur with a small sample sizes?	Over/or Underestimation.	
	In which way would known biases associated with the respective study design affect the reliability of the data?	Already discussed above.	
Representativeness	Does the sampling strategy and the size of the sample ensure representativeness for the study population?	1./4. Yes, the number of the participants in the VELS study is 732 which are not breastfed. 2. The number of aluminium measurements ranges from 21 – 500.	0 0
	Can results of the sample be transferred to the target population and the scope (regional, temporal) of the exposure assessment?	1. The VELS study was performed in 2001/2002 and consumption habits may have changed. 1. Moreover, the consumption within the two 3-day-diaries dietary intake were taken as representative for the whole year.	(--/++) (++)
	Which assumptions and extrapolations are made, described and justified?	2. Aluminium measurements have been conducted between 2002-2015 but aluminium content may have changed.	?-/+
		3. Assumptions for cocoa content in	-/+

		chocolate cream were made (10%).	
Details of correla-tions/dependencies	Have relevant correlations between influencing factors (e.g. consumption and body weight) been described and taken into account in the model (e.g. in-take/breathing rate/body surface per kg body weight)? If there are correlations and structural dependencies, were they described in a transparent and logical way?	1./4. The correlation between consumption and bodyweight was accounted for. 1. Seasonality of chocolate consumption was not accounted for. Yes	0 (++) Rated under scenarios 0
Evaluation methodology	**With deterministic estimates:** Are the statistical descriptions reported in a transparent and logical manner? Is the sample large enough to estimate the required parameters with sufficient precision? Which level of statistical precision (standard error of estimate, SEE and confidence intervals) has the exposure estimate? **With probabilistic estimates:** Are the statistical methods and selection criteria for distributions described in a transparent and logical way? Were considerations reported or additional data sets used to justify the selection of the distribution type? Is the sample size for the parameters considered large enough to accommodate the required distribution, especially extreme percentiles, with sufficient precision? Was the precision of the distribution fit and the corresponding parameters specified by providing confidence intervals, goodness-of-fit measures (e.g. Kolmogorov-Smirnov distance)? Were relevant statistical indicators (e.g. skewness, mean/median ratio,	Yes 1./4. The number of the participants in the VELS study which are not breastfed is 732. 2. The number of aluminium measurements ranges from 21 – 500. Not estimated. Not applicable	0 -/+ --/++

	percentiles) of the empirical and the parametrically distribution compared and dis-cussed? Which assumptions were made to fit a distribution using small samples? What are the consequences of these assumptions for the target variable of the exposure assessment?		

5.2.4.7. Question list for qualitative uncertainty analysis in relation to the procedure for exposure calculation

Table 23: Question list in relation to the documentation of the exposure calculation

Criteria	Questions	Uncertainty assessment	Direction
Deviations	Are there deviations between the exposure model and the implementation of the calculation method?	No, the initial exposure model was implemented in SPSS.	0
Review of cal-culations	Are there potential sources of error in the technical realisation of the model calculation process, the applied algorithms, the programming process (e.g. incomplete documentation, reproducibility) or the input of controlling variables (e.g. selection of the random number generator, number of iterations)?	No	0
Deficient report compilation	Are there potential sources of error in the report compilation process?	No	0
Verification	Were the units in the calculation controlled (e.g. within the SI system)?	Yes, the units were controlled.	0
	Was the model implementation independently repeated or assessed for quality?	Yes, it is a standard model for exposure assessments regarding consumption of food.	0

5.2.4.8. Standardised qualitative presentation of the findings of uncertainty analysis for primary documentation

The following table summarizes the findings of the uncertainty analysis. The symbols used where defined in Table 16 (see BfR guideline p.37 (Heinemeyer *et al.*, 2015)):

Table 24: Standardised presentation of the findings of uncertainty analysis

Identified aspects and magnitude of uncertainties in the exposure assessment					
	Degree of uncertainty	**Confidence in the knowledge base**			**Subjectivity of choices**
1. Goal and question formulation of the exposure assessment					
Question formulation	? -/+				High
Context	? -/+				
Protection perspective	0				
Protected population	-/+				Medium
Goals of protection	0				
Protection level	--/++				High
Restriction of scope	---				
2. Exposure scenario					
Characterisation of the noxious agent	0				
Exposure source and origin, exposure routes and pathways (media)	---/+++ ? -/+	Low			
Possible exposure paths	? -/+	Low			
Exposed groups of people/population	0				
Exposure events	--				
Assumed spatial, time-based and situational differences/lifestyles/modes of behaviour and microenvironment	--/++	Medium			
Risk management measures	Not appl.				
3. Exposure model					
Exposure estimator: definition of the target variable	-/+				
Concept and assumptions used for the translation of the scenario into model equations	0				
Dependencies/Correlations	0				
Model structure, e.g. stratifications	0				
Choice of model equation	--/++				Medium
Model extrapolation	++				
Risk management measures	Not appl.				
4. Parameters (to be completed separately for each parameter)					
	1. Consumed amount of foods	2. Aluminium content	3. Cocoa content of chocolate cream	4. Body weight	

Expert opinions, default assumptions	0	0	0	0		
Definition, units and quantification of the influencing variables	++	0	0	0		
Reliability of measurements	--/++	-/+	-/+	0		
Quality of the data sources	0	0	0	0		
Study population	0	0	Not appl.	0		
Representativeness	++ --/++	--/++ ? -/+	-/+	0	Medium|Low| High|High	
Correlation structure	0	Not appl.	Not appl.	0		
Evaluation methodology	-/+	--/++	Not appl.	-/+		
Method of calculation						
Deviations	0					
Review of calculations	0					
Deficient report compilation	0					
Verification	0					

5.2.4.9. Conclusion of the uncertainty analysis using the BfR guidance

The uncertainty analysis has shown that many sources of uncertainties are prevalent. The dominant uncertainties are:

1) It is not known how aluminium enters into cocoa/chocolate products. One effect is that it cannot be determined whether aluminium content is brand dependent or not. As a result, it can neither be confirmed nor excluded that some consumers may eat products with high aluminium content over a long period of time.
2) Consumption data was taken from the VELS study, which was conducted 2001/2002. Consumption habits may have changed during this time with unclear magnitude.
3) Aluminium measurements were taken between 2002 and 2015. Under consideration of the first point, it cannot be assessed whether and to which degree aluminium content of chocolate/cocoa products have changed until 2017.
4) Products like mixed-milk beverages containing cocoa, chocolate bars, muffins containing chocolate, and chocolate cakes were not included in the initial assessment. Therefore, the aluminium intake might be significantly underestimated.
5) Since the 6-days dietary diaries of the VELS survey were extrapolated to determine long-term consumption, the variability (caused by intrapersonal day-to-day variation of consumption as well as by seasonality) will be overestimated. For the 95th percentile of population this will lead to an overestimation of aluminium intake.

5.3. Priorisation of uncertainties

In order to prioritise uncertainties, first all identified uncertainties (by the use of the BfR question lists) were listed. A simple sensitivity analysis was carried out in order to reach a rough understanding about the possible impact of these uncertainties. These results were used to divide the uncertainty analysis into parts.

5.3.1. List of uncertainties

Sources of uncertainties were identified using the BfR question list (Table 17-23). These uncertainties are summarised below:

1) It is not clear whether aluminium content is brand-dependent and to which degree consumers are brand-loyal. This uncertainty can be directly linked to the missing knowledge of how aluminium enters into cocoa and chocolate products.
2) Dietary food study VELS was conducted in 2001/2002, but exposure assessment estimates exposure in 2017. Consumption habits of cocoa and chocolate may have changed during this time.
3) Coca content in chocolate-, nougat-, and cocoa cream is unknown.
4) Products like mixed-milk beverages containing cocoa, chocolate bars, muffins containing chocolate, and chocolate cakes were not included in the initial assessment.
5) Sampling error for the 95th percentile of the protection group, since only 732 participants (which are not breastfed) in VELS.
6) In order to derive the long-term intake of chocolate and cocoa, the 6 days dietary diaries are extrapolated. This may lead to overestimation for heavy consumers, since effects of intrapersonal variability (daily consumption varies from day to day) inflates the variability, which in turn leads to overestimation (increasing variability increases the value of the 95th percentile). Moreover, seasonality (chocolate is consumed more in winter than in summer) may lead to overestimation, since large consumption values in the VELS studies might be partly due to both three days dietary diary periods occurring in winter).
7) Errors in body weight determination.
8) Errors for the reported food weights.
9) Aggregation of different products to product groups.
10) Samples for aluminium measurements were taken before 2017.
11) Sampling of products for aluminium measurements is not necessarily representative for consumption.
12) Measurement error of aluminium content.
13) Sampling error for the aluminium content measurements.
14) Some aluminium measurements were below the limit of quantification.

5.3.2. Simple sensitivity analysis

In order to prioritise the sources of uncertainty, a simple sensitivity analysis is carried out. Only one source of uncertainty is varied for each time ("one at a time approach") and it was assumed that there is no dependency among the sources of uncertainties.

5.3.2.1. Preliminary Analysis to estimate the impact of each product group on intake of aluminium via chocolate and cocoa

Since the overall aim is to derive the aluminium intake of the 95th percentile of the population, the relative impact of different product groups on aluminium intake needs to be estimated. Especially for the 95th percentile, the task is non-trivial: different participants (within the VELS study) have different

ratios for the particular product groups. Additionally, if aluminium content of a product group or the consumed amount of the product group changes, the participant representing the 95th percentile may change too. Therefore, as a rough approach, for each product group the amount consumed was changed by ±10% and the appropriate 95th percentile of Al intake (assuming average Al concentrations) estimated. We have i=1, 2, …,8 different product groups (see Table 25). Let x_i denote the amount of the food consumed of the i*th* product group and P95(a x_i) the 95th percentile with given food amounts (in VELS) except product group i, where the original amount is multiplied with a factor a. Then we calculated (P95(1.1 x_i)-P95(0.9 x_i))/2 to derive a kind of rough partial derivative, which gives an approximate estimate about how much the Al intake of the 95th percentile changes if the amount of the consumed product group under consideration changes by 10%. That the amount of food per product group is varied by a rather large step (10%) is due to the fact that for very small changes only the composition of product groups of a single consumer would play a role, while a large step includes a broader number of heavy consumers. It should be noted though, that this is a very coarse approach that relies on interpolation as well as extrapolation. Also the real values can differ if consumption of more than one product group changes. But as a result, relative changes of consumption of a product group can be translated to an increase of aluminium intake for the 95th percentile (assuming average aluminium concentrations). The results are given in Table 25:

Table 25: Impact of relative changes of consumption of different product groups on intake of aluminium via chocolate and cocoa

Product group	Change of 95th percentile of aluminium intake if amount of consumed product group changes by 10% (in µg/kg per day)
Sugar panned chocolate	0.00
Milk chocolate/baking chocolate	0.55
Chocolate icing/ chocolate sprinkles/ chocolate coating	0.19
Chocolate with fillings	0.57
Dark chocolate	0.20
Cocoa powder	1.36
Beverage powder containing cocoa powder	1.97
Chocolate-, nougat-, and cocoa-cream	0.32

The Table 25 is used as follows: If e.g. consumption of dark chocolate is believed to be 15% higher than indicated by VELS data, than the increase in aluminium intake per day for the 95th percentile is estimated to $\frac{15\%}{10\%} * 0.2\ \frac{\mu g}{kg} = 0.3\ \frac{\mu g}{kg}$.

For children age 0.5 until 5 years, the product groups:

- Milk chocolate/baking chocolate,
- Chocolate with fillings,
- Cocoa powder,
- Cocoa containing beverage powder,

are most influential on the change of aluminium intake (regarding the 95th percentile of target population). For comparison, for heavy consumers (95th percentile) and average aluminium concentrations, the daily Al intake was 56 µg per kg bodyweight per day or around 390 µg per kg bodyweight per week.

5.3.2.2. Simple sensitivity analysis – detailed investigations

For each considered source of uncertainty assumptions were made for allowing a first estimate of their impact on the total aluminium intake. If necessary, Table 25 was used to estimate the impact of e.g. an increase of consumption of a certain product group or a change in aluminium content per product group. The result of the estimation is expressed as percentage of the original aluminium intake estimate of the 95th percentile (heavy consumer), which is given above - 56 µg per kg bodyweight per day. Afterwards the estimated impacts are compared with each other in order to prioritise the uncertainties.

(1) Aluminium content brand dependent/brand independent

If it is assumed that

- Aluminium content is perfectly brand dependent,
- high consumer are perfectly brand-loyal,
- consume throughout all product groups only products with high-content aluminium (95% percentile of aluminium content assumed),

then calculation yield a weekly amount of roughly 830 ng/kg aluminium intake, or an increase of more than 100% with respect to the value derived by mean aluminium content. Nevertheless, the combination of assumptions leads to a strong overestimation (especially that only high-contaminated products through all product groups are consumed). Therefore, the actual impact might be significantly smaller.

(2) Consumption trends since the completion of the VELS survey

The VELS survey was conducted between 2001 and 2002, but aluminium intake needs to be estimated for 2017. For some product groups, data could be found that allows comparing consumption in the past with the present. But the data is not accurate enough for an actual exposure estimate as it:

- describes consumption for the general population and not for children younger than 5 years
- for many of the used sources the study approach is not clear
- is unclear how heavy consumers are affected.

(2.1) Consumption of chocolate and chocolate containing products:

The consumption data are taken from Statista, a private provider for market and consumption data and relates to the whole population (Statista, 2018a). The following data were used:

2000: 8.25 kg per person per year

2005: 8.79 kg per person per year

2017: 9.48 kg per person per year (preliminary estimate)

It can be used to get a rough estimate of the relative change in chocolate consumptions. Using the rather strong assumption that the trend in consumption behaviour for the target population is the same as for the population as the whole it gives an idea of how large the change can be. In order to derive an estimate for 2001/2002 a linear trend between 2000 and 2005 is assumed and leads to a value of 8.41 kg for the time VELS was conducted. Compared with the value obtained in 2017, an increase of 13% can be stated (9.48/8.41=1.13). Using Table 25, an increase of 10% of chocolate related products (sum of the first 5 product groups in Table 25, which cover the chocolate products) would lead roughly to an increase of Al-intake of 1.5 µg per kg body weight per day. Therefore, an increase of 13% would yield an additional Al-intake of 2 µg per kg body weight per day or an overall impact of approximately 3%. It is unclear if this over- or underestimates the actual change.

(2.2) Consumption of cocoa powder

The data for cocoa powder were taken from Statista, a private provider for market and consumption data (Statista, 2018b)

2010: 0.72 kg per person per year

2017: 0.44 kg per person per year (estimation)

Using the consumption patterns for cocoa powder itself yields a decrease of 39% between 2010 and 2017. Two product groups are affected by cocoa powder: cocoa powder and beverage powder containing cocoa powder. A change of 10% for both of these product groups would change Al-intake by 3.3 µg per day per kg bodyweight (Table 25). A decrease of 39% therefore facilitates a decrease of roughly 13 µg per day per kg bodyweight or 23% of total aluminium intake for the 95th percentile of consumption. However, it is unclear whether this change also covers the difference between 2001/2 and 2017. Another way to estimate an overall change in consumption of cocoa powder would be the amount of consumed hot cocoa. Data exists for the change between 2003 and 2010 (no data was published before and after these years)and was once more obtained from Statista (Statista, 2018c):

2003: 3.4 % hot cocoa

2010: 2.9 % hot cocoa

This amounts to a reduction of 15 percent between these 7 years. Comparing this to the decrease in cocoa powder of 39% above it is unclear whether the figure above is sufficient to describe the total change between 2001/2 and 2017. The assumption with the largest impact on overall aluminium exposure would be to assume that the decline between a naive addition of the 15% decline with the 39% decline yields a total decline of almost 50% of cocoa powder consumption with an overall decrease of 30% of Al-intake for the 95th percentile. However the trend for cocoa powder and hot cocoa might as well be different with an unclear estimate on its impact.

(3) Cocoa content in chocolate-, nougat-, and cocoa-cream

In the original exposure assessment, it was mentioned that the cocoa content in chocolate-, nougat-, and cocoa-cream varied between 4-15% (10% was eventually chosen). This range corresponds to a

decrease of 60% and increase of 50% to the used value (of 10% cocoa content). Using Table 25, this product group accounts for 0.32 µg Al per day per kg bodyweight per 10% change. Therefore, if the upper value would have been chosen, this would result in 3% additional intake of Al. For the lower value it would result in 3% less intake. No additional assumptions were made as these figures were taken directly from the original assessment. Of course, products with a higher cocoa content might exist which might make this number slightly higher.

(4) Not considered products containing chocolate/cocoa

To estimate the potential additional intake from mixed-milk beverages containing cocoa the VELS data were analysed and yielded a daily intake of 4.15g per day per kg bodyweight for heavy consumers of this product group (95th percentile). As no concentration data were available in the original study an internet investigation for their cocoa content was performed and gave a range from 1.1-1.7% (product descriptions). See e.g.: https://www.codecheck.info/essen/milch_milchprodukte/milch_mischgetraenke_shakes.kat

Using a conservative assumption of 2% and assuming average Al-content for cocoa results in an additional Al intake of 12.6 µg/kg per day or additional 23% intake of Al for heavy consumers.

However the representativeness of the internet data is questionable. Additional beverages containing higher percentages of cocoa powder might exist and result in a higher impact. On the other hand, it is also highly questionable whether the current high consumers (regarding the product groups listed in Table 25) are also high consumers of mixed-milk beverages containing cocoa. Therefore the impact might be substantially smaller.

Other food items like muffins with chocolate or chocolate cake are not mentioned often in the VELS study, therefore there impact might be limited.

(5) Sampling error within the VELS study

In total, the number of participants (which are not breastfed) was N = 732 infants in the VELS study. For the 95th percentile, the variance VAR for the number of people with higher Al-intake per bodyweight can be calculated using the binomial distribution: VAR=N*p*(1-p) → VAR = 732*0.95*(1-0.95) ≈ 35. The relative standard error is therefore $\sqrt{35}/N = 0.008$. In turn, this means that the empirical 95th percentile has a relative standard error of 0.8%. Taking two standard errors, results that the empirical 95th percentile is actually between the 93.4% to 96.6 % percentile or 49-61 µg Al intake per day per kg bodyweight (using the appropriate percentiles of the VELS data with average aluminium concentrations) which translates to a maximal deviation of 13% for the Al-intake of heavy consumers (95th percentile).

(6) Variability

Intrapersonal variability (consumption varies for each person from day to day) is an effect that does not interfere with average consumption (taken over all participants of the VELS study). However, it artificially inflates the overall variability of the estimated long-term aluminium intake. Low or high consumption values might be partially due to chance, since participants may have eaten below or above average amounts of a certain food during the two three-day-diary periods of the VELS study. For the 95th percentile of population, this overestimated variability leads to an overestimation of aluminium intake. Moreover, effects of seasonality (chocolate may be consumed more in winter than in summer) might have a qualitatively similar effect - large consumption of chocolate within the VELS

study might rely on a disproportionally high fraction of participants with both three days dietary diary periods during the winter half year.

The influence of the variability on the results could not be estimated, but it will lead to an overestimation of long-term consumption for the 95th percentile of the population.

(7) Errors in body weight determination

Body weight for the VELS study was measured with an accuracy of 200 g. All scales were of the same type with the same precision. The lowest body weight of an individual listed in the VELS survey (not breastfeed) 6.4 kg. The worst case in terms of influence on the output of the exposure assessment would be a systematic bias affecting all weight measurements in the same direction (higher for an overestimation). The relative impact of 0.2 kg over 6.4 kg is 3% on total Al-intake. As the body weight is used multiplicative in the exposure calculation these 3% would also directly affect the result.

It should however be noted, that it is very unlikely that a measurement error will affect all measurements by the maximum precision in the same direction.

(8) Errors for the reported foods (and food weights)

Foods are measured with an accuracy of 1g. But for out-of-home consumption, the consumed amount is simply estimated. Such estimates can have a relative standard deviation from 50-100% (Ambrus et al. (2013)). Unfortunately, it is not possible to determine from the VELS data if the amount of food consumed was weighted or estimated.

(9) Aggregation of different products to product groups

Different products have been aggregated to product groups in order to ensure a representative number of aluminium measurements and a general compatibility with the data from the dietary study VELS. In the following, products within these product groups are treated as homogeneous, which is an assumption. However, the magnitude of this uncertainty is unknown. It may lead to an over/underestimation of the Al-intake for heavy consumers.

(10) Samples for aluminium measurements were taken before 2017

Aluminium measurements of chocolate and cocoa products were taken in the period of 2002-2015, but aluminium intake needs to be estimated for 2017. Neither the magnitude nor direction of this uncertainty is known.

(11) Sampling for aluminium measurements is not representative regarding consumption

The selection of product samples for aluminium measurements do not represent actual consumptions, therefore seldom consumed products might be overestimated. This may lead to an over/underestimation of the Al-intake for heavy consumers. The impact of this uncertainty is not known.

(12) Measurement error of aluminium content in chocolate and cocoa

The concentration measurements stem from routine inspections from Federal States Enforcement Agencies. Thus they are naturally rather heterogeneous and do have very different measurement precisions. The overwhelming majority of measured content data is more than 10-times larger than the limit of quantification and for many measurements this ratio is larger than 100.

Assuming that the precision of the measurement is – in the worst case - similar to the limit of quantification (LOQ) and using a very conservative estimate of the precision being 10% of the actual value (resulting in the LOQ being 10-times lower than the measurement) this would be the maximum impact on the aluminium intake.

It is very unlikely that this 10% will impact each value in the same direction, since it is a random error. Thus it was scaled with √(51) assuming equal chances of an upward or downward fluctuation for each measurement. 51 is the lowest number of measurements in a single food group for those that have high impact (see Table 25). The relative uncertainty from this source is thus 10%/√ (51) = 1.4%. Since other important food groups have much more individual data points and the accuracy is typically much higher, the 1.4% is likely an overestimation of the actual impact.

(13) Sampling error for the aluminium content measurements

To estimate the sampling error on mean concentration, one can look at the standard error for the measurements of the food groups. The relative standard errors are:

Sugar panned chocolate	5.6%
Milk chocolate/ baking chocolate:	4.8%
Chocolate icing/ chocolate sprinkles/ chocolate coating	14.5%
Chocolate with fillings	16.5%
Dark chocolate	2.3%
Cocoa powder:	1.7%
Beverages containing cocoa powder:	7.6%
Chocolate-, nougat-, and cocoa-cream	1.7%

These relative errors directly translate on the intake for each of the overall food groups. Propagating this to the overall Al-intake (using Table 25) yields the following result: 1.8 µg/kg bodyweight per day for one standard error. Using two standard errors, the relative impact could be 6% of Al intake for heavy consumer (95th percentile). If other than mean aluminium content values are used, the results will likely be different.

(14) Values of aluminium content measurements below the limit of quantification

From 1646 samples, 14 samples had measured aluminium values below the limit of quantification and were set 0. These are less than 1% of all samples, therefore the impact for aluminium intake is supposed to be even smaller.

5.3.2.3. Summary of simple sensitivity analysis and comparison with the BfR results of the qualitative assessment

The process of the simple sensitivity analysis allowed getting a rough impression of how sources of uncertainty may impact the result. However, it should be mentioned that not all present uncertainties

could be quantified here. Nevertheless, the findings indicate that the following uncertainties may significantly alter aluminium intake of the 95th percentile of the target population:

- Question whether aluminium content is brand dependent and consumers are brand-loyal or not
- Change of cocoa powder consumption since 2001/2002
- Not included food groups, especially ready-to-drink mixed milk beverages containing cocoa
- Sampling error regarding the 95th percentile of aluminium intake of the participants of the VELS survey as well as to a lower degree the sampling error of aluminium measurements.

In 5.2.4 a qualitative assessment of the uncertainties (employing the BfR guideline) has been carried out. It is interesting to compare these results with the ones of the simple sensitivity analysis here. While in general the results roughly correspond to each other, it should be noted that in the qualitative assessment, the importance of the sampling error regarding the 95th percentile of aluminium intake was not appropriately indicated. This underlines the advantage of a simple quantification using sensitivity analysis. However, the BfR guideline emphasises the importance of sensitivity analysis and therefore such procedure could in principle also been carried using the BfR uncertainty guideline.

5.4. Dividing the exposure assessment into parts

Since an exposure model exists, it seems wise to first decide which uncertainties can be implemented within this model. From the previous simple sensitivity analysis, it seems that the uncertainty whether aluminium content depends on brand in combination with brand-loyalty of consumers or not might be worth investigating in more detail. Therefore, two scenarios (and subsequently two model versions) are applied:

- Scenario 1: It is assumed that consumers eat chocolate and cocoa products without preference to only high or low aluminium content. Therefore, for the long-term intake, average aluminium content is assumed.
- Scenario 2: It is assumed that aluminium content is brand dependent as well as customers are completely brand-loyal within one product group, with no cross correlation to other product groups.

Moreover, it seems important (and easy to implement) to include ready-to-drink mixed milk beverages containing cocoa. The same approach as used for chocolate-, nougat-, and cocoa creme can be used, which requires an assumption of cocoa content of mixed milk beverages containing cocoa. Sampling error of aluminium measurements seems to be substantial and easily to be implemented for both scenarios. The same goes for the sampling error within the VELS study.

Furthermore, the performed simple sensitivity analysis indicated that changes in consumption of cocoa powder may result in a decrease of up to 30% of aluminium intake. However, doubts of the applicability of the data sources warrants caution. Therefore, the change in consumption of cocoa powder was selected as suitable topic for a single parameter expert knowledge elicitation (EKE).

The remaining uncertainties (plus new emerging uncertainties due to the model extensions) will be collectively assessed via an EKE in order to estimate overall uncertainty.

5.5. Characterising uncertainty for parts of the assessments and combining them

5.5.1. Model extensions

In order to help estimating aluminium intake, a probabilistic model was created. Nevertheless it still follows closely the initial model for the exposure assessment, using all product groups employed within this assessment with the addition of ready-to-drink milk beverages containing cocoa (where consumption data is also taken from the VELS study). For the product groups chocolate-, nougat-, and cocoa cream as well as ready-to-drink milk beverages containing cocoa, no direct measurements of aluminium concentration exist. As in the previous model, for chocolate-, nougat-, and cocoa cream a cocoa content of 10% was assumed. For the newly introduced product group ready-to-drink milk beverages containing cocoa a cocoa content of 2% was set. Then the model proceeded with aluminium content of cocoa powder. Two versions of the model were developed, each realizing a different scenario.

Scenario 1: It is assumed that consumers eat chocolate and cocoa products without preference to only high or low aluminium content. Therefore, for the long-term intake, average aluminium content is assumed.

Scenario 2: It is assumed that aluminium content is brand dependent as well as customers are completely brand-loyal within one product group, with no cross correlation to other product groups.

For scenario 1, the aluminium intake of each of the 732 participants of the VELS study is estimated by using average aluminium concentrations given already in Table 15.

For scenario 2, the aluminium content of the product groups given in Table 15 are described by parametric distributions. Distributions were chosen among the set of the following two-parametric, non-negative distributions: Weibull distribution, gamma distribution, log-normal distribution, log-logistic distribution as well as the inverse gamma distribution. Parameters were estimated using the method of moments in order to reproduce the empirical mean. The distribution with the smallest Anderson-Darling distance was chosen. The chosen distributions and the values for the parameters are given in Table 26. The fits can be seen in the Appendix.

Simulations for scenario 2 required that for each participant a random percentile for aluminium content for each product group is chosen (using the fitted two parameter distributions) and subsequently the aluminium intake was estimated. For the product groups chocolate-, nougat-, and cocoa cream as well as ready-to-drink milk beverages containing cocoa, the percentiles were chosen from the distribution of cocoa powder and combined with the respective assumptions regarding cocoa content (10% and 2% respectively).

Table 26: Fitted distributions for aluminium concentration for each product group

Product group	Distribution	Parameter 1	Parameter 2
Sugar panned chocolate	Log-logistic	a = 20.32	b = 3.295
Milk chocolate/baking chocolate	Gamma	k = 1.150	b = 0.04494
Chocolate icing/ chocolate sprinkles/ chocolate coating	Gamma	k = 2.280	b = 0.068778
Chocolate with fillings	Log-normal	μ = 2.668	δ = 0.9894
Dark chocolate	Gamma	k = 3.773	b = 0.07581

Cocoa powder	Gamma	k = 7.050	b = 0.04634
Beverage powder containing cocoa powder	Gamma	k = 3.414	b = 0.1016

Explanation to Table 26:

Gamma distribution: $f(x) = \frac{b^k}{\Gamma(k)} b^{k-1} e^{-bx}$

Log-normal distribution: $f(x) = \frac{1}{x\delta\sqrt{2\,\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\delta^2}}$

Log-logistic distribution: $f(x) = \frac{\frac{b}{a}\left(\frac{x}{a}\right)^{b-1}}{\left(1+\left(\frac{x}{a}\right)^b\right)^2}$

5.5.1.1. Considered sources of uncertainty in the model

Scenario1:

- Uncertainty of mean values for aluminium content for each product group

This uncertainty was assumed to be normally distributed; the normal distribution was parametrised using the empirical mean and standard error.

- Sampling error regarding the VELS study

For each Monte Carlo run, for all 732 participants of the VELS study the individual aluminium intake is calculated. From these 732 original values, in total 732 values are sampled (sampling with replacement) and finally, the 95th percentile estimated.

Scenario 2:

- Uncertainties for estimated parameters of the fitted distributions (aluminium content)

For each set of estimated parameters, 10,000 times samples were drawn (same sample size as the original aluminium measurements) from the respective distributions. For each of these bootstrap samples, parameters of the same distribution type were estimated (using the method of moments again) and saved. For the actual Monte Carlo simulation, for each run a set of parameters is randomly chosen from these 10,000 newly generated parameter sets.

- Sampling error regarding the VELS study

For each Monte Carlo run, for all 732 participants of the VELS study the individual aluminium intake is calculated. From these 732 original values, in total 732 values are sampled (sampling with replacement) and finally, the 95th percentile estimated.

5.5.1.2. Model results

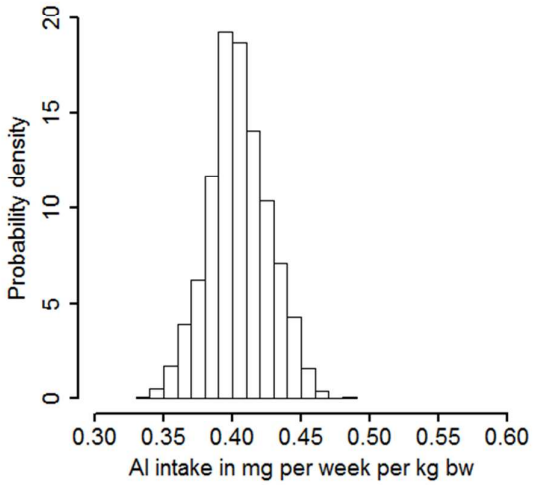

Figure 4: Aluminium intake for the 95th percentile of the target population for scenario 1

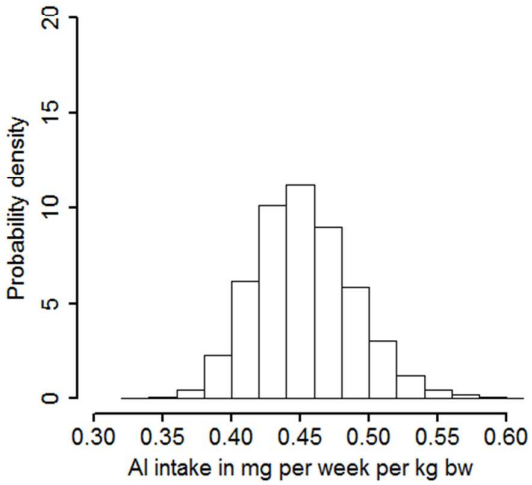

Figure 5: Aluminium intake for the 95th percentile of the target population for scenario 2

Table 27: Model results of aluminium intake for the 95th percentile of the target population for different percentiles of the resulting distribution

Percentile	1%	25%	50%	75%	99%
Scenario 1 in µg/(week*kg bw)	363	394	408	425	461
Scenario 2 in µg/(week*kg bw)	379	428	450	475	547

Surprisingly, for the 95th percentile of the population under consideration, the values for the quartiles are roughly only 9-12 % larger for scenario 2 than scenario 1. This rather modest influence of brand-loyalty might be due to the significant influence of several product groups, which mitigates the effect. However, it also should be noted that the distribution for scenario 2 displays a larger uncertainty, with a 50% larger inter-quartile distance than in scenario 1.

5.5.2. Expert knowledge elicitation (EKE) for changes of cocoa powder consumption

Because of the potential indicated by the simple sensitivity analysis, the change of cocoa powder consumption from 2001/2002 to 2017 was individually assessed by an EKE. The product groups cocoa powder and beverage powder containing cocoa powder are affected. EFSA has published a detailed guidance on EKE (EFSA, 2014); additional information can be found in an e-learning course by Tony

O'Hagan O'Hagan (O'Hagan). Throughout this project, the Sheffield method was employed, which emphasises that the experts should in general strive to find a consensus for their estimates. In total, 5 percentiles were elicited: the lower bound (1st percentile), the lower quartile (25th percentile), the median (50th percentile), the upper quartile (75th percentile) and the upper bound (99th percentile). The following people were involved:

Elicitator (Moderator):	Olaf Mosbach-Schulz (EFSA)
Experts:	Christine Sommerfeld (BfR) – expert for this case study (performed the original exposure assessment)
	Carolin Fechner (BfR) – expert on food supply chains
	Katrin Blume (BfR) – expert on consumption data
	Oliver Lindtner (BfR) - expert on exposure assessment
	Christian Jung (BfR) – expert on uncertainty, coworker in this case study
	Thomas Schendel (BfR) – expert on uncertainty, main investigator for the uncertainty analysis of the case study
Date:	07.05.2018

As outlined in the simple sensitivity analysis, there is a potential that cocoa powder consumption has been reduced by up to 50%. However during the EKE, the relevance of the respective sources has been doubted. First, no explanation could be thought of for such a decline, second it does not seem to be reflected in the variety of products containing cocoa powder in the supermarkets, and third it was judged that the sources are less relevant for young children, since they only describe consumption of overall population. Nevertheless, it was acknowledged that the resulting distribution should have a slight negative bias, attributing to the fact that there is no indication of increasing cocoa powder consumption which could counter the existing (weak) evidence of decreasing consumption. The following results were obtained:

Table 28: Results of EKE for changes of cocoa powder consumption from2001/2002 to 2017

Percentile	**1%**	**25%**	**50%**	**75%**	**99%**
Change of cocoa consumption in %	-30	-15	-5	7.5	20

Using the MATCH Uncertainty Elicitation (Morris *et al.*, 2014) tool to fit a distribution, the scaled beta distribution with parameters $\alpha=1.238$ and $\beta=1.177$ seemed to fit the elicited quartiles best.

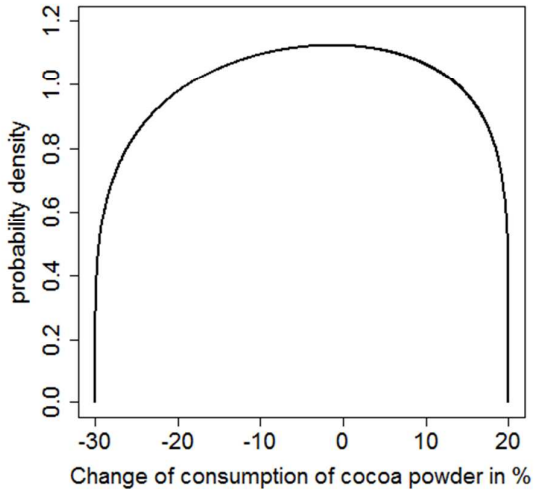

Figure 6: Elicited distribution for the change of consumption cocoa powder from 2001/2002 to 2017

5.5.3. Implementing the uncertainty of changes of cocoa powder consumption into the model

The results of the EKE regarding changes of cocoa powder consumption from 2001/2002 to 2017 need to be incorporated into the model. Therefore, all consumption values for cocoa powder and beverage powder containing cocoa powder were multiplied with a constant factor, sampled for each Monte Carlo run from the distribution obtained by the previous EKE. The results and comparison with the sole model results can be seen below:

Table 29: Scenario 1 – aluminium intake for 95th percentile of the target population

Percentile	1%	25%	50%	75%	99%
Scenario 1 in µg/(week*kg bw) initial model result	363	394	408	425	461
Scenario 1 in µg/(week*kg bw) after one-parameter EKE	339	377	401	427	472

Table 30: Scenario 2 – aluminium intake for 95th percentile of the target population

Percentile	1%	25%	50%	75%	99%
Scenario 2 in µg/(week*kg bw) initial model result	379	428	450	475	547
Scenario 2 in µg/(week*kg bw) after one-parameter EKE	363	416	443	472	551

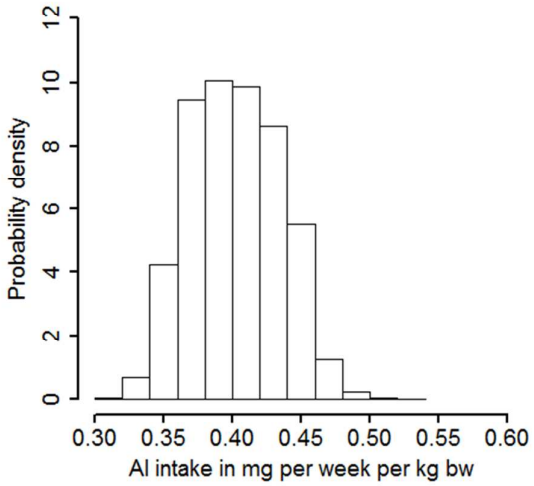

Figure 7: Aluminium intake for the 95th percentile of the target population for scenario 1

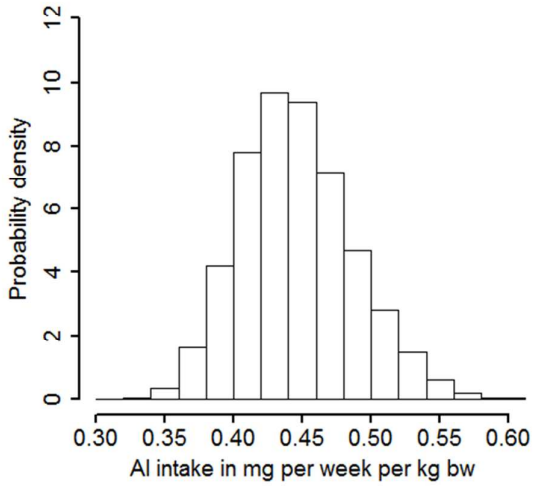

Figure 8: Aluminium intake for the 95th percentile of the target population for scenario 2

It can be seen that in both scenarios a minor shift to smaller values of aluminium intake occurs but that especially for scenario 1, the width of the distribution increases noticeably. The latter effect is not that pronounced for scenario 2, attributing to the fact that the original width of the distribution was already larger before the uncertainty of cocoa powder consumption changes have been incorporated.

5.6. Characterising overall uncertainty

The following list of uncertainties was collectively assessed during an EKE in order to quantify the overall uncertainty. The uncertainties arising from further model assumptions were added too.

- Change of consumption habits from 2001/2002 to 2017 (excluding cocoa powder)
- Coca content in chocolate-, nougat-, and cocoa cream
- Cocoa content in ready-to-drink milk beverages containing cocoa
- Foods containing cocoa/chocolate not taken into account in the model
- Uncertainty arising from the extrapolation of VELS data (6 days) to derive long-term intake
- Errors in body weight determination
- Errors for the reported food weights
- Aggregation of different products to product groups
- Samples for aluminium measurements were taken before 2017
- Sampling of products for aluminium measurements is not necessarily representative for consumption
- Measurement error of aluminium content and values below the limit of quantification
- Modelling error I: Finite number of Monte Carlo simulations

- Modelling error II: Implementation of changes of cocoa powder consumption as homogeneous factor for all participants of the VELS survey
- Modelling error III: For scenario 1, assumption that empirical mean (of aluminium measurements) is normally distributed around the true mean
- Modelling error IV: For scenario 2, assumption that aluminium measurements per product group can be described by two-parametric distribution

First, the EKE the quantification of the overall uncertainty of scenario 1 has been carried out. The same procedure as described in 5.5.2 was performed.

Elicitator (Moderator):	Olaf Mosbach-Schulz (EFSA)
Experts:	Christine Sommerfeld (BfR) – expert for this case study (performed the original exposure assessment)
	Carolin Fechner (BfR) – expert on food supply chains
	Oliver Lindtner (BfR) - expert on exposure assessment
	Christian Jung (BfR) – expert on uncertainty, coworker in this case study
	Thomas Schendel (BfR) – expert on uncertainty, main investigator for the uncertainty analysis of the case study
Date:	07.05.2018

To assess so many uncertainties collectively has been rated to be difficult and therefore this insecurity has been translated into a wide range of the elicited distribution. Regarding bias, different opinions were voiced. On the one side, the effect of overestimation due to the extrapolation of the 6-days diaries to long-term consumption for the 95th percentile was judged by one expert to have a major impact on the result. Other experts challenged this argument, believing that other uncertainties, especially the non-inclusion of not incorporated foods like cake containing chocolate may easily offset this effect. As a conclusion, no major bias was incorporated into the elicited distribution. As result of the EKE, the following quantiles were elicited (other values given for comparison):

Table 31: Scenario 1 – aluminium intake for 95th percentile of the target population

Percentile	**1%**	**25%**	**50%**	**75%**	**99%**
Scenario 1 in µg/(week*kg bw) initial model result	363	394	408	425	461
Scenario 1 in µg/(week*kg bw) after one-parameter EKE	339	377	401	427	472
Scenario 1 in µg/(week*kg bw) after overall EKE	200	320	400	470	600

The next step was the execution EKE for the overall quantification of uncertainty for scenario 2.

Elicitator (Moderator): Carolin Fechner (BfR)

Experts: Christine Sommerfeld (BfR) – expert for this case study (performed the original exposure assessment)

Katrin Blume – expert on consumption data

Christian Jung (BfR) – expert on uncertainty, coworker in this case study

Thomas Schendel (BfR) – expert on uncertainty, main investigator for the uncertainty analysis of the case study

Date: 09.05.2018

The main reasons voiced in the previous EKE have been repeated. Moreover, it was acknowledged that for scenario 2 the distribution is already broader than for scenario 1; additionally the errors due to the fitting of aluminium content per product group with two parameter distributions will lead to additional uncertainty. Therefore it was judged that the distribution should be broader than for scenario 1. The elicted quartiles are given as follows:

Table 32 Scenario 2 – aluminium intake for 95th percentile of the target population

Percentile	**1%**	**25%**	**50%**	**75%**	**99%**
Scenario 2 in µg/(week*kg bw) initial model result	379	428	450	475	547
Scenario 2 in µg/(week*kg bw) after one-parameter EKE	363	416	443	472	551
Scenario 2 in µg/(week*kg bw) after overall EKE	200	350	440	550	700

Using the MATCH Uncertainty Elicitation Tool (Morris *et al.*, 2014), for both scenarios the scaled beta distribution seemed to fit the elicited quartiles best. In case of scenario 1, the parameters are α=1.693 and β=1.735 and in case of scenario 2, the parameters are α=1.500 and β=1.542.

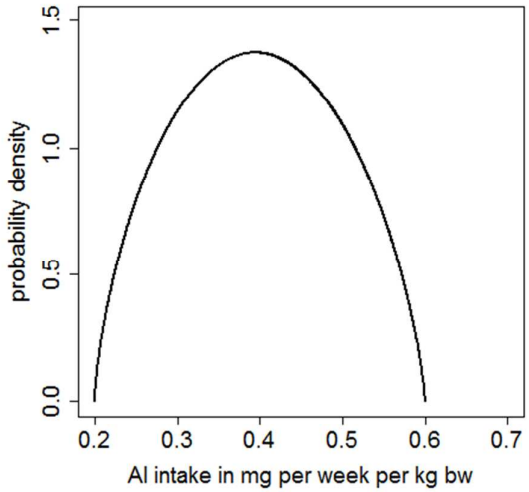

Figure 9: Elicited distribution for aluminium for the 95th percentile of the target population for scenario 1

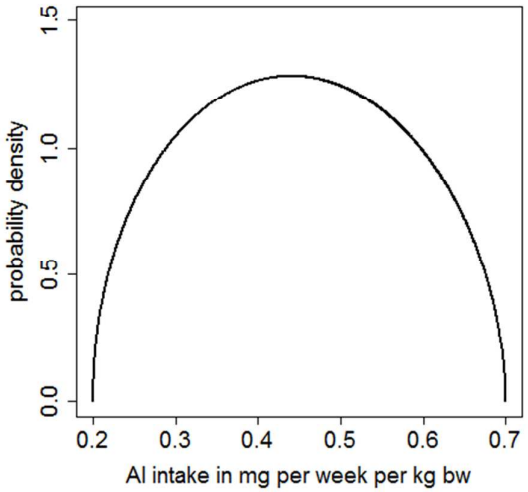

Figure 10: Elicited distribution for aluminium for the 95th percentile of the target population for scenario 2

5.7. Discussion of results of the uncertainty analysis for the exposure assessment of aluminium in cocoa and chocolate

The uncertainty analysis for aluminium intake in 2017 for the 95th percentile of children age 0.5-5 years which are not breastfed supported the conclusion of the initial BfR exposure assessment: a relevant part of the tolerable weekly intake of aluminium is exhausted by the consumption of cocoa and chocolate.

Assuming mean aluminium concentrations per product group, the initial BfR exposure assessment derived a value of 347 µg/(week*kg bw) (for the 95th percentile of population), while the final result of this uncertainty analysis yielded 400 µg/(week*kg bw) for the median of scenario 1 (interquartile range 320-470 µg/(week*kg bw)). The slightly larger value (for the median) reflects among other things the inclusion of a further product group.

The results for assuming brand-dependence of aluminium content and brand loyalty were obtained by different approaches in the initial BfR exposure assessment and this uncertainty analysis. In the initial BfR exposure assessment, average aluminium content was assumed except a selected product group, where the 95th percentile of aluminum content was chosen. For this percentile, only consumers of this specific product group were considered. In opposite, in this uncertainty the 95th percentile refers to all participants of the VELS study. Furthermore, for each participant a value for aluminium content of each product group was randomly sampled. Therefore, it is not surprising that, depending on the selected product group, the initial BfR assessment yielded values of 488-565 µg/(week*kg bw), while here a lower value of 440 µg/(week*kg bw) for the median was derived (interquartile range 350-550 µg/(week*kg bw)). In recognition of the differences between both approaches, the findings seem credible.

Nevertheless, the results of this uncertainty analysis have revealed large uncertainties: for scenario 1, the interquartile range is nearly 38% of the median value; for scenario 2, this value reaches approximately 45%.

Surprisingly, the pure model results (before considering any results from expert elicitation) indicate that brand-loyalty (in combination with brand-dependence of aluminium content) only yields 10 % larger results for the median of the 95th percentile of the population than the use of average aluminium content. To understand this effect, it is helpful to recall that they are in total nine product groups. While some product groups do not contribute much to aluminium intake for the 95th percentile of the population, the fact remains that it is unlikely that within the Monte Carlo simulations an individual would have large aluminium content values for all relevant product groups. Moreover, other sources of uncertainty play a role in determining the aluminium intake of the 95th percentile of the population; therefore the effect of above average aluminium concentrations caused by brand-loyalty further diminishes.

6. Conclusions and Recommendations

6.1. Conclusions

Within this project, the uncertainty documents of EFSA and BfR were compared. As an overall conclusion, it can be stated that both documents establish a framework for uncertainty analysis, which is inspired by the same philosophy. One important difference between both guidelines is the scope: the BfR guideline is only applicable to exposure assessments, while the EFSA guidance includes all risk assessment within EFSA mandate. The BfR guideline on uncertainty contains extensive question lists to identify sources of uncertainties for exposure assessments. It enables the risk assessor to find uncertainties along the complete path of the exposure assessment: starting at the assessment question, continuing with the exposure scenario, choice of model and its parameters, and finally investigating the method of exposure calculation. In contrast, the EFSA guidance gives only little practical advice on how to identify uncertainties (while it still requires the risk assessor to list all uncertainties).

Another significant difference is that the BfR describes in detail one qualitative method to assess the uncertainties. Quantitative uncertainty analysis is mentioned, but the guideline does not elaborate further on practical implementation. On the other side, EFSA presents a toolbox of qualitative and quantitative methods which are introduced, advantages- and disadvantages discussed and an application of the method showcased. Moreover, EFSA recommends a quantification of the overall uncertainty.

Two case studies were chosen to apply and compare both uncertainty guidelines. Both topics dealt with the uncertainty analysis of exposure assessments, in order to be in the scope of the respective uncertainty documents of EFSA and BfR. For the first case study, an exposure assessment regarding ARS/SC in eel was chosen. The initial BfR report concluded that a risk assessment is not feasible, due to the lack of reliable data regarding toxicity of ARS/SC and not sufficient data on eel consumption. The case study is a good example for situations with poor data and little available time (only 1-2 weeks for the initial assessment). An important characteristic of this case study is that originally no exposure model was developed. This had significant implications on the performed uncertainty analyses.

The uncertainty analysis following the BfR guideline led to an extensive set of identified sources of uncertainties. As a result, the missing knowledge deficiencies were reported, which could be used as reference for future investigations. Moreover, the results also serve as a good foundation for a quantitative uncertainty analysis, since the identification and qualitative assessment of the uncertainties facilitate the planning of the uncertainty assessment. In this regard, the uncertainty guideline of the BfR can be seen as complementary to the one of EFSA. However, it should be mentioned that it was found to be difficult to differentiate the degree of impact (number of plusses/minuses) in the qualitative assessment.

For the uncertainty analysis partly employing EFSA's guidance document, it was decided to use a probabilistic model. For this case study, it should be noted that due to the lack of data, the analysis rested on many assumptions which in turn needed to be treated as scenarios. Nevertheless, the remaining uncertainties were successfully quantified. It should be noted that the EFSA guidance recommends quantifying the overall uncertainty, which was due to organisational reasons not done here. Nevertheless, if a toxicological threshold value for ARS/SC could be established, it would either be much larger than the overwhelming part of the probability distribution of eel marker intake per bodyweight, indicating that health risks from the consumption of marked eel by ARS/SC are not to be expected. Or a non-negligible part of the probability distribution would be larger than the threshold value, indicating the need for further refinement. Following the philosophy of EFSA guidance document, expert knowledge elicitations could be used to narrow down the large uncertainties

assumed for the ARS/SC intake per eel and the fraction of ARS/SC accumulated in the edible eel tissue.

Due to familiarity with numerical approaches, the project team has chosen an approach based on an explicit simple mathematic model. It is very possible that alternative approaches that are more consistent with the EFSA guidance and that do not require a mathematical model would have been applicable within shorter time. Specifically, the EFSA guidance recommends for urgent assessments the use of a single expert knowledge elicitation (EKE) to assess the overall uncertainty. Without reference value, such a task is hard to solve, but not impossible. For the given example of eel marker intake, one could assume fish consumption to represent eel consumption and assume that all consumed eel contents 100% eel marker in order to establish an upper boundary. Even if such coarse estimates might not necessarily be helpful for the risk manager (which nevertheless can be the case), it illustrates what is possible given few resources and tight deadlines and may emphasise the need for further investigations. Moreover, it was noticed during the case study that a simplification of the BfR question list could further reduce the time required to follow this qualitative approach in emergency situations.

For the second case study, an exposure assessment of aluminium in cocoa and chocolate was chosen. The original BfR exposure assessment concluded that a significant part of the tolerable weekly intake of aluminium is exhausted by the consumption of products containing cocoa/chocolate, especially young children. The original assessment was characterised by sufficient data and more available time compared to the first case study. The uncertainty analysis presented here focused on the 95th percentile of children age 0.5 - 5 years, which are not breastfed. By applying the BfR guideline uncertainties were identified and qualitatively assessed. Examples for the discovered uncertainties are the missing knowledge on how aluminium enters into cocoa/chocolate products (which also affects whether brand-loyalty plays a role or not), the aged consumption data, the non-inclusion of several foods containing cocoa or chocolate, and the extrapolation of the short-term intake estimated by the VELS study to long-term intake. The latter would overestimate the variance and therefore overestimate the aluminium intake for the 95th percentile of the population.

Subsequently the quantitative assessment of these uncertainties was performed using the EFSA guidance building on the list of identified uncertainties with the BfR guideline. In particular, a simple sensitivity analysis was carried out in order to prioritise the uncertainties. As a result it was decided to employ a probabilistic model in two versions: one describing a scenario with average aluminium concentrations and another one that assumes brand-dependence of aluminium content and brand-loyalty within a single product group. Sampling errors for both, consumption data and aluminium content were also incorporated in this model as well as consumption data for ready-to-drink mixed milk beverages containing cocoa. Apart from the model, the change of cocoa powder consumption from 2001/2002 to 2017 was estimated using an expert knowledge elicitation (EKE). The result indicated a large uncertainty with an overall slightly decreasing tendency. This elicited parameter was incorporated into the model.

As a last step, in order to characterise the overall uncertainty, another EKE was performed for both scenarios outlined above. All uncertainties which had not been considered previously were assessed collectively. While the location of the prior obtained probability distribution did not change much, its variance increased remarkably, indicating the additional uncertainty of the added sources of uncertainty.

The utilisation of the EKE has been a new experience for the participating experts. It was widely acknowledged that it is a very valuable tool to stimulate discussions which in turn allowed the participants to consider new angles of the topic at hand. On the other hand, the experts felt (especially for the overall characterisation of uncertainty) that using quantitative judgements is rather difficult and subjective. However, the EKE is a method in order to express the belief of an expert in quantitative terms, which is subjective by definition. In order to simplify the complexity that was

encountered during the EKE to quantify the overall uncertainty, more uncertainties could have been assessed individually before. In turn, this would have required more time.

The uncertainty analysis for the second case study has shown that the uncertainty guidances of BfR and EFSA work very well together. The identification of uncertainties by the BfR question lists and the quantitative assessment of these uncertainties using the EFSA guidance (including the simple sensitivity analysis for prioritisation of the uncertainties, dividing the assessment into parts, estimating the uncertainties for each part, combining the uncertainties, and characterising the overall uncertainty) led to a successful quantitative description of uncertainty for aluminium intake via cocoa and chocolate which reinforces the conclusion of the initial BfR exposure assessment and gives detailed information of the range of uncertainty.

As a main result of the two case studies, it could be shown that both guidances complement each other very well. The BfR guideline on uncertainty excels mainly in the identification of sources of uncertainties. Moreover, the qualitative assessment of the BfR guideline could also be used as a preselection measure to determine those uncertainties which might be assessed in more detail via a simple sensitivity analysis later on. On the other side, the EFSA guidance document provides a complete framework of how to conduct a quantitative uncertainty analysis along with a description of many different quantitative methods. During the project, it was recognized that the BfR guideline only describes a qualitative approach for assessing the uncertainties in detail. This procedure is necessarily subjective. It was found to be difficult to differentiate the degree of impact (number of plusses/minuses) in the qualitative assessment. The BfR acknowledges that a quantitative uncertainty analysis can be employed as refinement of the uncertainty analysis, but it does not elaborate on how to conduct it. A more refined description for quantitative uncertainty analysis is therefore missing.

The use of expert knowledge elicitation (EKE) in order to quantify uncertainty has been applied for the second case study twice. First, a single parameter was elicited via an EKE and second, an EKE was employed to assess the overall uncertainty of the exposure assessment. Especially the latter use of the EKE – the assessment of the overall uncertainty at the end of the uncertainty analysis is a crucial part of the EFSA guidance. In the second case study presented here, a couple of complex problems needed to be assessed in the overall EKE. Not only a wide range from different uncertainties needed to be considered, but also statistical effects needed to be accounted for. For example, how does the width of an existing probability distribution (as a proxy for variance) influences the overall uncertainty distribution if more sources of uncertainties are added? How do the uncertainties affect the 95th percentile of population? It is hard to account for such questions accurately in an EKE, or to phrase this issue differently: it is questionable whether for 50% of all EKEs used to quantify the overall uncertainty, the interquartile range would include the true value. But the decisive point is a different one – the quantification of uncertainty (often only possible by an EKE) provides valuable information to risk managers that would not be available otherwise. While the approach to quantify the overall uncertainty via an EKE might be approximate, it will lead to better results (and be more helpful to risk managers) than if the overall assessment of uncertainties would not be conducted.

To summarise, the project has shown that uncertainty analysis is a topic of crucial importance. Both, the BfR and EFSA guidance, have their unique approach in carrying out such analyses, which nonetheless are very compatible with each other.

6.2. Recommendations

The comparison of both uncertainty guidelines together with their application to two case studies has yielded valuable experience. As a result, recommendations to both guideline documents can be given. During this project, it has become apparent that the BfR guideline document needs to be updated.

- The guideline does not include any case study for illustration purpose yet. But this would be very helpful for people with no prior experience in uncertainty analysis in understanding how to carry out an uncertainty analysis by their own.
- While the BfR guideline mentions quantitative uncertainty analysis, it does not elaborate on this topic, leaving the reader without any guidance for this matter. In contrast, the EFSA guidance lists many methods (qualitative and especially quantitative) which are introduced, advantages- and disadvantages discussed and an application of the method showcased. It might be worthwhile to reference this toolbox of methods to use in case a quantitative uncertainty analysis is required.
- While using the question lists to identify sources of uncertainty, it became evident that many redundancies regarding the question lists for identifying uncertainties exist. The same sources of uncertainty can often occur in both, the scenario and the parameter section. This can make the use of the BfR guideline rather tedious. Formulating a more straightforward version would improve the situation.

Recommendations for the EFSA guidance are related to the fact, that comparison of both uncertainty documents has shown that the BfR guideline excels in the identification of uncertainties using question lists guiding the risk assessor through all elements of an exposure. In contrast, the EFSA guidance gives only little practical advice on how to identify uncertainties and leaves it to the risk assessor to care for the complete listing of uncertainties.

- It should be kept in mind that the scope of the EFSA guidance covers the full risk assessment process and is therefore much broader than the scope of BfR guidance focusing on exposure assessment only. Therefore, the use of a common question list for identifying uncertainties for all four steps of the risk assessment (hazard identification, hazard characterisation, exposure assessment and risk charcterisation) for all scientific assessments within EFSA scope would hardly be feasible. However, it would be possible and is recommended by the authors that for each topic (like biological hazards, chemical contaminants, pesticides …) separate question lists are created to support the identification of the respective sources of uncertainty.
- Another issue deals with the expert knowledge elicitations (EKEs) recommended by EFSA for characterising the overall uncertainty. As shown in the second case study, the experts may have to deal with very complex issues all at once. Hence, the elicited results are necessarily only approximate (valuable nonetheless for decision makers). Therefore, the communication of such results should avoid overconfidence to reflect this fact. In this regard, it should be also emphasised that in general, for large uncertainties (like in the first case study) it is better to report only ranges without central estimates, since their informative value diminishes in the face of large uncertainties and may lead to inappropriate anchoring.

Finally, it has become evident that there are many open questions with regard to uncertainty. One important topic would be the use of EKE.

- One question that needs further investigation is: Are the results of EKEs reproducible? In order to study this question, one could conduct several EKEs for the same topic (possibly also stratify whether few or large amounts of relevant knowledge is available) and compare the results.
- More difficult would it be to determine whether the setting of EKE is able to avoid cognitive biases known to be present in individual decisions involving statistics (see e.g. (Kahnemann *et al.*, 1982)), hence to avoid systematic bias in the results of EKEs. An experimental approach pursuing this research would need to find questions that can be unambiguously answered but which are nonetheless not available to the invited experts.

- Another important issue concerns the role of uncertainty analysis in risk management. How much more useful are the results of quantitative compared to qualitative uncertainty analysis for risk management? How did the communication of uncertainties (qualitative and/or quantitative) actually changed risk management measures? A research project tackling such question would need to engage deeply with risks managers to obtain reliable results. But these results may help risk assessors to understand better which kind of information is important in risk management and subsequently adjust uncertainty analysis and their communication accordingly.

References

Ambrus Á, Horváth Z, Farkas Z, Dorogházi E, Cseh J, Petrova S, Dimitrov P, Duleva V, Rangelova L, Chikova-Iscener E, vaskainen ML, Pakkala H, Heinemeyer G, Lindtner O, Schweter A, Trichopoulou A, Naska A, Sekuła W, Guiomar S, Lopes C, Torres D (2013). Pilot study in the view of a Pan-European dietary survey – adolescents, adults and elderly. *EFSA Supporting Publications, 10* (11). doi: *10.2903/sp.efsa.2013.EN-508*

ANSES FAfF, Environmental and Occupational Health & Safety (2016). Prise en compte de l'incertitude en évaluation des risques: revue de la littérature et recommandations pour l'Anses.

Banasiak U, Heseker H, Sieke C, Sommerfeld C, Vohmann C (2005). Abschätzung der Aufnahme von Pflanzenschutzmittel-Rückständen in der Nahrung mit neuen Verzehrsmengen für Kinder [Estimation of the dietary intake of pesticide residues based on new consumption data for children]. *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz, 48* (1): 84–98. doi: *10.1007/s00103-004-0949-6*

Benford D, Halldorsson T, Jeger MJ, Knutsen HK, More S, Naegeli H, Noteborn H, Ockleford C, Ricci A, Rychen G, Schlatter JR, Silano V, Solecki R, Turck D, Younes M, Craig P, Hart A, Von Goetz N, Koutsoumanis K, Mortensen A, Ossendorp B, Germini A, Martino L, Merten C, Mosbach-Schulz O, Smith A, Hardy A (2018). The principles and methods behind EFSA's Guidance on Uncertainty Analysis in Scientific Assessment. *EFSA Journal, 16* (1). doi: *10.2903/j.efsa.2018.5122*

BfR (2017a). *Stellungnahme zur Aalmarkierung mit Alizarin-Rot S und Strontiumchlorid.* Report.

BfR (2017b). *Sachstand zum Thema Aluminium in Kakao und Schokolade*. Report.

Caraguel JM, Charrier F, Mazel V, Feunteun E (2015). Mass marking of stocked European glass eels (Anguilla anguilla) with alizarin red S. *Ecol Freshw Fish, 24* (3): 435-442. doi: *DOI 10.1111/eff.12158*

Codex (2015). *Codex Alimentarius Commission. Procedural Manual*. 23. ed.

Durif C, Dufour S, Elie P (2005). The silvering process of Anguilla anguilla: a new classification from the yellow resident to the silver migrating stage. *Journal of Fish Biology, 66* (4): 1025-1043. doi: *10.1111/j.0022-1112.2005.00662.x*

ECHA, European Chemicals Agency (2012). Guidance on information requirements and chemical safety assessment Chapter R.19: Uncertainty analysis.

EFSA-AFC, European Food Safety Authority (2008). Scientific Opinion of the Panel on Food Additives, Flavourings, Processing Aids and Food Contact Materials on a request from the European Commission on the safety of aluminium from dietary intake. *The EFSA Journal, 6* (7): 754 *(122 pp.)*. doi: *10.2903/j.efsa.2008.754*

EFSA (2014). Guidance on Expert Knowledge Elicitation in Food and Feed Safety Risk Assessment. *EFSA Journal, 12* (6). doi: *10.2903/j.efsa.2014.3734*

EFSA (2016). Guidance on Uncertainty in EFSA Scientific Assessment [Draft]. EFSA Journal, 273 pages.

EFSA (2017). Guidance on Uncertainty in EFSA Scientific Assessment [Second Draft]. EFSA Journal, 37 pages.

EFSA, European Food Safety Authority (2012). *When Food Is Cooking Up a Storm – Proven Recipes for Risk Communications*.

Ehlscheid N, Lindtner O, Berg K, Blume K, Sommerfeld C, Heinemeyer G (2014). Selten verzehrte Lebensmittel in der Risikobewertung. Ergebnisse einer Telefonbefragung in Deutschland. *Proceedings of the German Nutrition Society*.

Heinemeyer G, Mosbach-Schulz O, Kreienbrock L, Schümann M, Filter M, Greiner M, Herzler M, Lindtner O, Kurzenhäuser S, Röder B (2015). Guidance document on uncertainty analysis in exposure assessment. Recommendation of the Committee for Exposure Assessment and Standardisation of the Federal Institute for Risk Assessment (BfR). Series: *BfR-Wissenschaft*, Vol. *03/2015*, 56 pp. Berlin, DE: Bundesinstitut für Risikobewertung (BfR). Retrieved from http://www.bfr.bund.de/cm/350/guidelines-on-uncertainty-analysis-in-exposure-assessments.pdf – accessed: 2015-10-26.

Heseker H, Oeppining A, Vohmann C (2003). *Verzehrsstudie zur Ermittlung der Lebensmittelaufnahme von Säuglingen und Kleinkindern für die Abschätzung eines akuten Toxizitätsrisikos durch Rückstände von Pflanzenschutzmitteln (VELS)*. Report: *Forschungsbericht für das Bundesministerium für Verbraucherschutz, Ernährung und Landwirtschaft*, III+71 pages. Paderborn DE: Universität Paderborn.

IPCS, International Programme on Chemical Safety & IOMC, Inter–Organization Programme for the Sound Management of Chemicals (2008). Uncertainty and Data Quality in Exposure Assessment. Part 1: Guidance Document on Characterizing and Communicating Uncertainty in Exposure Assessment. Part 2: Hallmarks of Data Quality in Chemical Exposure Assessment. Series: *IPCS Harmonization Project Documents*, Vol. *6*, 175 pp. Geneva, CH: World Health Organisation (WHO). Retrieved from http://www.who.int/ipcs/methods/harmonization/areas/hazard_assessment/en/ – accessed: 2015–09–01.

Jacoby D & Gollock M (2014). *Anguilla anguilla. The IUCN Red List of Threatened Species 2014:e.T60344A45833138*. Website: http://dx.doi.org/10.2305/IUCN.UK.2014-1.RLTS.T60344A45833138.en – accessed: 05-10-2017.

Kahnemann D, Slovic P, Tversky A (1982). *Judgement under uncertainty: Heuristics and Biases*. Cambridge University Press, Cambridge UK.

Kroes R, Dentonkelaar E, Minderhoud A, Speijers G, Vonkvisser D, Berkvens J, Vanesch G (1977). Short-term toxicity of strontium chloride in rats. *Toxicology, 7* (1): 11-21. doi: *10.1016/0300-483x(77)90033-6*

Kullmann B, Neukamm R, Thiel R (2017). Mass-marking of farmed European eels (Anguilla anguilla (Linnaeus, 1758)) with alizarin red S. *Journal of Applied Ichthyology, 33* (5): 914-917. doi: *10.1111/jai.13424*

Morris DE, Oakley JE, Crowe JA (2014). A web-based tool for eliciting probability distributions from experts. *Environmental Modelling & Software, 52*: 1-4. doi: *10.1016/j.envsoft.2013.10.010*

MRI (2008). *Nationale Verzehrsstudie II – Die bundesweite Befragung zur Ernährung von Jugendlichen und Erwachsenen. Ergebnisbericht, Teil 1 & Ergänzungsband: Ausgewählte Ergebnisse nach Schichtindex*. Report, xxvi+144 & 20 pages. Karlsruhe DE: Max Rubner-Institut – Bundesforschungsinstitut für Ernährung und Lebensmittel. Retrieved from https://www.mri.bund.de/fileadmin/MRI/Institute/EV/NVS_II_Abschlussbericht_Teil_1_mit_Ergaenzungsbericht.pdf – accessed: 2013-05-28.

Neukamm R (2009). *FIAF-Pilotprojekt zur Förderung des Aales in den Gewässersystemen Nord-Ostsee-Kanal und Elbe-Lübeck-Kanal*. Vol. 2017. Report. Retrieved from http://angelgewaesser.lsfv-sh.de/images/downloads/Jahresbericht_2006-2008_FIAF.pdf – accessed: 11-10-2017.

Neukamm R & Hempel M (2017). *Der Einfluss von Aalbesatzmaßnahmen auf die Bestandsstruktur in den Gewässersystemen Nord-Ostsee-Kanal und Elbe-Lübeck-Kanal. Zwischenbericht 2016*. Report. Retrieved from https://www.schleswig-

holstein.de/DE/Fachinhalte/F/fischerei/Downloads/berichtAalbesatzmassnahmen.pdf?__blob=publicationFile&v=2 – accessed: 11-10-2017.

O'Hagan T. *E-learning course: Probabilitstic Judgements for Expert Elicitation*. Website: http://www.tonyohagan.co.uk/shelf/ecourse.html – accessed: 18-06-2018.

Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, Tarantola S (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications, 181* (2): 259-270. doi: *10.1016/j.cpc.2009.09.018*

Simon J (2007). Age, growth, and condition of European eel (Anguilla anguilla) from six lakes in the River Havel system (Germany). *Ices Journal of Marine Science, 64* (7): 1414-1422. doi: *DOI 10.1093/icesjms/fsm093*

Simon J, Dorner H, Richter C (2009). Growth and mortality of European glass eel Anguilla anguilla marked with oxytetracycline and alizarin red. *J Fish Biol, 74* (1): 289-295. doi: *10.1111/j.1095-8649.2008.02117.x*

Statista (2018a). *Pro-Kopf-Konsum von Schokoladenwaren in Deutschland in den Jahren 1970 bis 2017 (in Kilogramm)*. Report. Retrieved from https://de.statista.com/statistik/daten/studie/72632/umfrage/pro-kopf-verbrauch-von-schokoladenwaren-in-deutschland/ – accessed: 17-05-2018.

Statista (2018b). *Absatz im Markt von Kakao*. Website: https://de.statista.com/outlook/30030000/137/kakao/deutschland.

Statista (2018c). Verteilung des Konsums von Heißgetränken in Deutschland nach Art in den Jahren 2003 bis 2010. Retrieved from https://de.statista.com/statistik/daten/studie/225311/umfrage/verteilung-des-konsums-von-heissgetraenken-in-deutschland-nach-der-art/.

Stein F, Doering-Arjes P, Fladung E, Brämick U, Bendall B, Schröder B (2016). Downstream Migration of the European Eel (Anguilla Anguilla) in the Elbe River, Germany: Movement Patterns and the Potential Impact of Environmental Factors. *River Research and Applications, 32* (4): 666-676. doi: *10.1002/rra.2881*

WHO-IPCS, World Health Organisation - International Programme on Chemical Safety (2004). IPCS Risk Assessment Terminology - Part 2: IPCS Glossary of Key Exposure Assessment Terminology.

Wickstrom H & Sjoberg NB (2014). Traceability of stocked eels - the Swedish approach. *Ecology of Freshwater Fish, 23* (1): 33-39. doi: *10.1111/eff.12053*

Appendix A – Fit of aluminium content for each product group

For seven product groups, direct aluminium measurements are available. For scenario 2, these aluminium measurements were fitted with two-parametric distributions. The quality of these fits can be seen in the following figures.

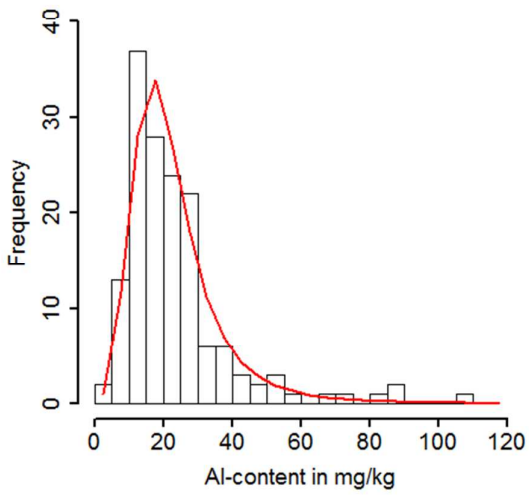

Figure 11: Sugar panned chocolate

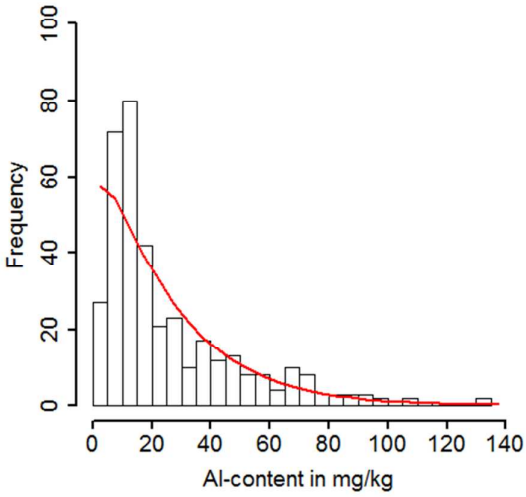

Figure 12: Milk chocolate/baking chocolate

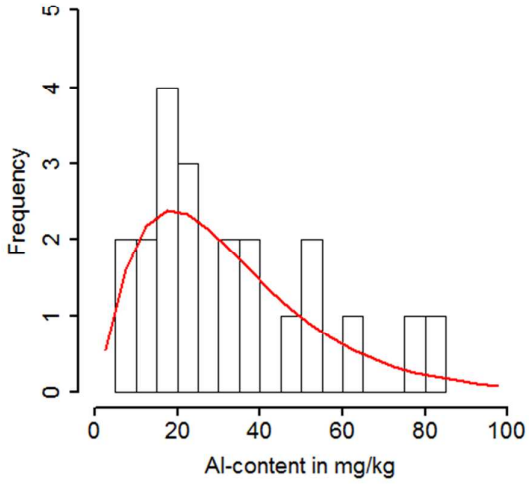

Figure 13: Chocolate icing/ chocolate sprinkles/ chocolate coating/sprinkles

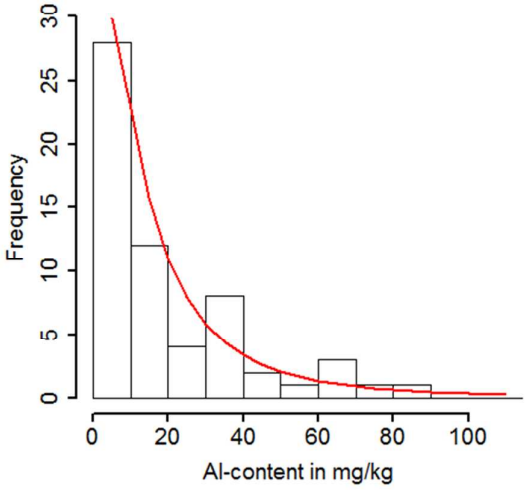

Figure 14: Chocolate with fillings

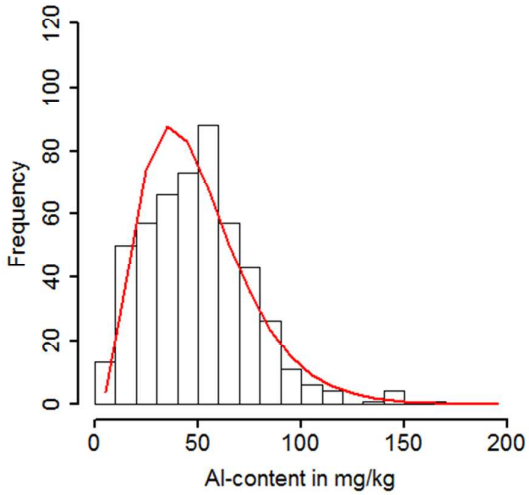

Figure 15: Dark chocolate

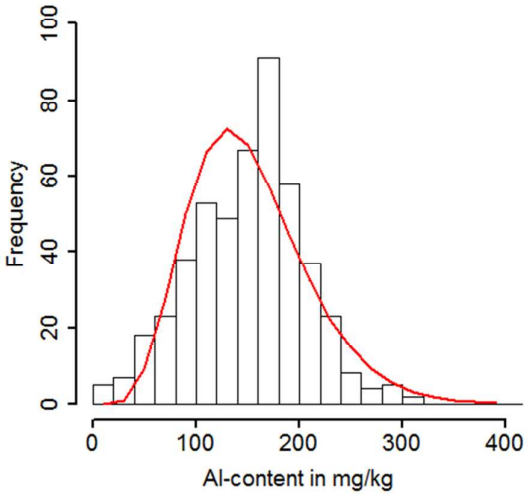

Figure 16: Cocoa powder

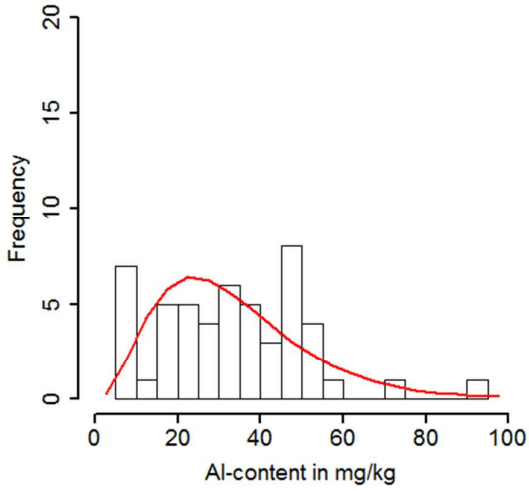

Figure 17: Beverage powder containing cocoa powder