External Scientific Report

APPROVED: 19 July 2024
doi: 10.2903/sp.efsa.2024. EN-9008

# Proposal for a qualification system for New Approach Methodologies (NAMs) in the food and feed sector: example of implementation for nanomaterial risk assessment

Andrea Haase[1], João Barroso[2], Alessia Bogni[2], Susanne Bremer-Hoffmann[2], Valerie Fessard[3], Arno C. Gutleb[4], Jan Mast[5], Emily McVey[6], Birgit Mertens[5], Agnes G. Oomen[6,7], Vera Ritz[1], Tommaso Serchi[4], Katherina Siewert[1], Deborah Stanco[2], Shirin M. Usmani[1], Eveline Verleysen[5], Olimpia Vincentini[8], Meike van der Zande[9], Francesco Cubadda[8]

[1]German Federal Institute for Risk Assessment (BfR), Berlin, Germany; [2]European Commission, Joint Research Centre (JRC), Ispra, Italy; [3]French Agency for Food, Environmental and Occupational Health & Safety (ANSES); [4]Luxembourg Institute of Science and Technology (LIST); [5]Sciensano, Brussels, Belgium; [6]National Institute for Public Health and the Environment (RIVM), Bilthoven, the Netherlands; [7]University of Amsterdam, Amsterdam, the Netherlands; [8]Istituto Superiore di Sanità - National Institute of Health (ISS), Rome, Italy; [9]Wageningen Food Safety Research, part of Wageningen University and Research (WFSR), Wageningen, the Netherlands

## Abstract

Plenty of new approach methodologies (NAMs) for risk assessment have been developed but only some are included in OECD Test Guidelines (TGs) for regulatory implementation. Nevertheless, NAMs are increasingly applied, e.g. for nanomaterial (NM) risk assessments. The EFSA Guidance on NM risk assessment suggests that NAM-derived data concerning degradation/dissolution (in relevant biofluids), intestinal uptake/crossing, genotoxicity, cytotoxicity, oxidative stress, (pro-)inflammatory potential and barrier integrity, for many of which no OECD TGs exist, have to be evaluated first. Consequently, NM risk assessments involve data from non-guideline studies, requiring time-consuming and challenging case-by-case evaluations. Establishing an OECD TG is a formal process aiming for international use according to the Mutual Acceptance of Data (MAD). However, not every promising NAM can be prioritised for OECD TGs. A qualification, based on an expert opinion, may enable an efficient use of adequate NAMs for a specific context-of-use. Furthermore, it supports the optimisation of promising NAMs for regulatory applications. Existing qualification systems

operate in the context of e.g., drug development tools (FDA) and research and development into pharmaceuticals (EMA). The NAMS4NANO consortium was tasked to propose a generic framework for a qualification system for chemical risk assessment in the food and feed sector to speed up the regulatory use of NAMs. Here we describe our proposal including the process and evaluation criteria. A detailed test method description, preferably as standard operating procedures (SOPs), describing the set-up of the NAM including its application and evaluation phase is crucial. Furthermore, the scientific validity, i.e. its reliability and relevance for the context-of-use, needs to be demonstrated, for which we suggest a less rigorous process compared to OECD TGs. We propose to initially establish a qualification system for NM risk assessment, aligned with the EFSA framework. This document is an interim version to stipulate a broader discussion among experts and stakeholders.



**Keywords:** qualification, validation, regulatory implementation, new approach methodologies (NAMs), nanomaterials, risk assessment

**Question number:** EFSA-Q-2024-00291

**Correspondence:** MESE@efsa.europa.eu

**This document is an interim version, which describes an initial proposal for a qualification system addressing NAMs for nanomaterials (NMs) with the intention to stipulate a broader discussion among experts and stakeholders. The document is open for feedback and comments until 31 December 2024.**

**Acknowledgements:** We would like to pay our special regards to the EFSA Methodology and Scientific Support (MESE) Unit for their support of the NAMS4NANO Project and their help with this interim report on a NAMs qualification system, in particular to Claudia Roncancio-Pena, Maria Chiara Astuto, Irene Cattaneo, and Adriana Scattareggia Marchese. Moreover, we are very grateful for all the comments and helpful suggestions, which we received from EFSA experts and external advisors. In particular, we would like to thank the EFSA WG on NAMs: Qasim Chaudhry, Maria Bastos, Jacqueline Castenmiller, Roland Franz, Henriqueta Louro, Alicja Mortensen, Anil Patri (US FDA), Elijah J. Petersen (US NIST), Michael Empl (EMA), Camelia Constantin (ECHA) and Marianne Matzke (ECHA). Special thanks also to Maurice Whelan and the EURL ECVAM team for useful comments and discussions. Finally, we acknowledge the useful information provided by additional experts within our institutions. Specifically, we acknowledge Michael Oelgeschläger (BfR) for providing useful insights into ongoing OECD activities and Philip Marx-Stoelting (BfR) for informing us on currently ongoing discussions on method readiness evaluation within the European Partnership for the Assessment of Risks from Chemicals (PARC).

**Suggested citation:** Haase A, Barroso J, Bogni A, Bremer-HoffmannS, Fessard V, Gutleb AC, Mast J, McVey E, Mertens B, Oomen AG, Ritz V, Serchi T, Siewert K, Stanco D, Usmani SM, Verleysen E, Vincentini O, van der Zande M, Cubadda F, 2024. Proposal for a fit for purpose qualification system for New Approach Methodologies (NAMs) in the food and feed sector. EFSA supporting publication 2024:EN-9008. 96 pp. doi:10.2903/sp.efsa.2024.EN-9008

**ISSN:** 2397-8325

# Summary

Huge investments in research projects have led to the emergence of many promising new approach methodologies (NAMs) to support risk assessment. However, only a tiny fraction is validated and established as Organization for Economic Cooperation and Development (OECD) Test Guidelines (TGs) for implementation in the regulatory context. It is widely acknowledged that many NAMs have unrealised potential for application in chemical risk assessment beyond prioritisation and screening. The comprehensive analysis of the reasons for this "valley of death" where NAMs fail to move beyond their establishment into validation and regulatory acceptance remains outside the scope of this document. Briefly, several publications (Westmoreland et al., 2022; Schmeisser et al., 2023; Gourmelon et al., 2024; Groenewold et al., 2024) have identified issues related to regulatory implementation of NAMs namely, i) unclear regulatory applications and uncertainties on relevance, ii) incomplete SOP(s), iii) unclear/ incomplete data-interpretation procedures, iv) unclear applicability domain and iv) lack of quality systems. These issues fall within the overarching challenge that there is insufficient support for validation. Despite missing validation and lagging establishment of OECD TGs, many NAMs are becoming widely established as tools in hazard identification, hazard characterisation and/or risk assessment of innovative consumer products by industries. Consequently, data from non-guideline studies, which are often complex and time-consuming to assess, is increasingly being submitted to regulators. The lack of harmonisation renders risk assessments in such cases very challenging.

Nanotechnology is an area in which the implementation of NAM-based approaches in the risk assessment process is promising and already established by the EFSA Guidance on risk assessment of NMs, in the following abbreviated as EFSA Guidance on Nano-RA (EFSA Scientific Committee, 2021a). NAMs are foreseen in the first steps of the risk assessment process to address specific nanoscale considerations, minimising the need for additional animal studies. However, for NMs most NAMs (with only a few exceptions) are not yet validated (Usmani et al., 2024). Additional challenges render method validation much more time and resource intense. For instance, NMs require a much more comprehensive physicochemical characterisation (not only of the pristine materials). Moreover, issues such as dispersion stability, dosimetry, agglomeration, dissolution, or assay interferences have to be tackled, rendering validation studies more complex. Data from non-guideline studies are routinely included in NM risk assessment, demanding for a case-by-case decision for each submission.

Method validation is required to establish OECD TGs that fall under the Mutual Acceptance of Data (MAD) of OECD Member Countries, facilitating the international use of test methods. However, not every promising NAM can be prioritised for inclusion in an OECD TG. In such cases, qualification, which is based on an expert opinion on the adequacy of a method for a well-defined use in a specific context (context-of-use concept), can be a promising solution. As a consequence, a NAM can be applied for the "qualified use" without the need to re-evaluate the method case-by-case in each application. Qualification also requires a demonstration of the reliability and relevance of the method but the process can be less rigorous due to the specific context-of-use. Moreover, qualification is better suited to support the optimisation of promising NAMs.

Methodology

Firstly, existing qualification systems that operate in the context of drug development tools (FDA, 2020) and research and development into pharmaceuticals (EMA, 2008) served as examples to prototype a fit-for-purpose framework in the food and feed sector. However, as the existing systems qualify NAMs for drug development tools, additional considerations are needed for NAMs for the safety assessment in the food and feed sector. We identified NAMs for NMs as a promising area to explore such a system. Therefore, we closely follow the EFSA Guidance on Nano-RA (EFSA Scientific Committee, 2021a). Furthermore, OECD GD 211 was identified as a key publication for describing non-guideline methods to facilitate their regulatory application. Finally, we identified three key approaches for assessing method readiness to support method validation, namely Bal-Price et al. (2018) being established for developmental neurotoxicity (DNT), the Pepper framework (Crouzet et al., 2023) being established for Endocrine Disruptors (ED) and the TRAAC framework (Shandilya et al., 2023) being established for nano-specific methods. We carefully compared all these documents with respect to the underlying criteria and how they are organized into sections/ sub-sections in order to propose a suitable approach for evaluating the readiness of NAMs in our qualification system for NMs.

Results

We propose a generic framework for a NAMs qualification system in the food and feed sector. The method submitter (i.e. developer, applicant or sponsor of the method) can submit a NAM, which is sufficiently developed for evaluation to an EFSA expert panel. The qualification system follows a sequential framework starting with a letter of intent (LoI) as an initial submission, which is evaluated to make a decision which route to follow, a fast track directly to qualification or a sequential approach via an interim submission (to seek for qualification advice) to the final submission of a full qualification proposal (to seek for a qualification opinion). The qualification system itself can be regarded as a practical tool, which allows regulators to a) evaluate NAMs regarding their adequacy for a specific context-of-use, b) support the optimisation of submitted NAMs and c) ultimately facilitate their regulatory implementation for risk assessments of NMs in the food and feed sector. It covers all phases of the process, namely i) the submission phase, ii) the evaluation phase and iii) the outcome phase. We propose a clear structure for the process, the individual steps and we provide guidance for each of the phases. This guidance explains i) how to structure the submissions, ii) which type of information and evidence is required, iii) what will be the criteria for the evaluation, iv) what is the expected duration of the process, v) what are the possible outcomes and vi) what are the consequences of a particular outcome/ decision.

We recommend to initially establish the system for NAMs for NMs aligned with the EFSA Guidance on Nano-RA (EFSA Scientific Committee, 2021a). We propose to initially establish a qualification programme "NAMs for nano", covering NAMs for a) NM physicochemical characterisation, b) characterisation of NM in relevant biological fluids (e.g., to assess solubility, dissolution/degradation, other relevant particle transformations that have an impact on the assessments); c) toxicity screening (specifically involving intestinal uptake/crossing, genotoxicity, cytotoxicity, reactivity/ oxidative stress, (pro-)inflammatory

responses and barrier impairment). The criteria to evaluate method readiness have been structured into three overarching sections, namely i) detailed test method description (preferably as SOPs) covering three parts (set-up of the NAM, its application and evaluation phase), ii) relevance and iii) reliability in the context of the regulatory application. In addition, there is a section covering essential general information that is not intended for evaluation but only for providing a summary and overview. It is important to note that all criteria will be evaluated together such that the order in which they are mentioned does not imply a prioritisation.

Importantly, we have already initially tested the evaluation criteria for one selected example, the triculture system (Vincentini et al., 2022), which can be applied to investigate NM uptake/ transport across intestinal barrier and to evaluate NM effects on barrier integrity.

Conclusions and Recommendations

We propose a fit-for-purpose qualification system to be applied for NAMs in the food and feed sector, which shall be initially established for NMs. We suggest to first establish three selected programmes and also propose some guidance through a criteria catalogue for evaluating these methods.

We recommend to firstly test the system to evaluate selected NAMs in the context of NM risk assessment case studies such as those currently ongoing in our consortium within Lot 2. This will allow for a critical discussion of the proposed system, the suggested process and the proposed criteria. Therefore, at this stage this report is an interim version only, which should mainly serve as a basis for a discussion with EFSA, Member States and relevant stakeholders. The document is open for public consultation. Together, this will help to fine-tune and finalise the approach.

# Table of contents

# 1 Background

## 1.1 Terms of Reference as provided by the requestor

This document was produced in the context of the EFSA NAMS4NANO Project Lot 1 with the aim to prototype a qualification system for facilitating the regulatory use of NAM tools and methods in the food and feed sector.

The title and number of this grant is NAMS4NANO - Integration of New Approach Methodologies results in chemical risk assessments: Case studies addressing nanoscale considerations (GP/EFSA/MESE/2022/01)[1].

The first call of the EFSA NAMS4NANO Project on the 'Integration of New Approach Methodologies results in chemical risk assessments: Case studies addressing nanoscale considerations' was subdivided into three Lots:

- Lot 1: Review of tools and developing a 'Qualification System for NAMs';
- Lot 2: Risk assessment case studies;
- Lot 3: Methodological and generic case studies.

The Lot 1 of this grant was awarded by EFSA to a consortium of organisations lead by the German Federal Institute for Risk Assessment (BfR). Co-beneficiaries are (i) Istituto Superiore di Sanità (ISS), the Italian National Institute of Health, (ii) the French Agency for Food, Environmental and Occupational Health & Safety (ANSES), (iii) Sciensano, (iv) the Dutch National Institute for Public Health and the Environment (RIVM), (v) Wageningen Food Safety Research, part of Wageningen University and Research (WFSR), and (vi) the Luxembourg Institute of Science and Technology (LIST). Fraunhofer ITEM is involved as a sub-contractor of the BfR. Singapore Food Agency (SFA) is included as an additional international partner. Furthermore, the Joint Research Centre (JRC) of the European Commission is participating in the project as a partner.

## 1.2 Additional information

The same consortium was awarded the three grants and cooperation is maintained between the different Lots for the duration of the project to ensure methodological consistency and maximise the results from the activities.

[1]https://www.efsa.europa.eu/en/art36grants/article36/gpefsamese202201-nams4nano-integration-new-approach-methodologies-results

## 1.3 Introduction

### NAMs and their role in risk assessment in the food and feed sector

To date, animal tests remain standard practice for the risk assessment of chemical substances but the establishment, validation and regulatory implementation of alternative methodologies, often collectively referred to as New Approach Methodologies (NAMs), have made great progress. Our consortium recently provided an overview on NAMs with a specific focus on those being relevant for nanomaterials (NMs). In this context we proposed the following NAMs working definition (Usmani et al., 2024):

> Any technology, methodology or approach that can provide useful information to support chemical risk assessment by informing on the hazard (considering toxicodynamics and toxicokinetics) and exposure of a chemical without the use of animal tests shall be considered a New Approach Methodology (NAM). This includes *in silico*, *in chemico*, *in vitro, ex vivo* and specifically for NMs, also physicochemical characterisation approaches. Any non-animal method can qualify as a NAM when it is validated or when it has been demonstrated to be scientifically valid for application in risk assessment.
>
> In most cases various NAMs have to be applied in combination, usually in frameworks describing tiered approaches, to respond to questions of regulatory relevance. However, any approach modifying an existing animal test would not be considered a NAM *per se*, even though innovative methods might be applied for this purpose. In line with this, the application of simple whole organisms *in vitro* models would qualify as NAMs as long as they can provide useful information for risk assessment and/or enable better design of *in vivo* testing. Animal tests might be included in such frameworks along with NAMs, usually in the higher tiers.

Importantly, the advantages of NAMs go beyond ethical considerations. NAMs can provide faster readouts in higher throughput in comparison to animal tests which are time and resource demanding low-throughput assays. Moreover, NAMs offer flexibility with options to combine various assays, test different cell models in parallel and/or multiplex readouts. Thus, they can be more efficient and informative than animal tests. Critically, issues regarding interspecies concordance and extrapolation from experimental animals to humans can be avoided. Notably, the measured pathological and behavioural endpoints in animal studies rarely provide meaningful mechanistic insights on the observed effects. In contrast, NAMs allow for unravelling toxicity mechanisms in a relatively direct manner relevant to human health adverse outcomes. Importantly, NAMs as testing tools are integral to keep pace with innovation, for the steadily increasing number of new chemicals and novel materials, or to address new areas of concern such as mixture toxicity. In addition, NAMs can be highly beneficial for the specific consideration of vulnerable subgroups of the general population. Therefore, NAMs show potential to significantly advance hazard and risk assessment in the future.

However, these potential gains from NAM-based approaches cannot be realised until their associated regulatory challenges are resolved. Unfortunately, their regulatory implementation

remains limited and is lagging behind their establishment for several reasons. Some of the major general obstacles are:

(i) The slow pace in standardisation and validation of NAMs;
(ii) The lack of harmonised reporting templates for data and meta-data with some notable exceptions (e.g., the OECD Harmonised Template (OHT) 201);
(iii) The limited extent to which data interpretation is harmonised;
(iv) The wide variety of biological models with different levels of complexity, each with specific advantages and limitations, which is also hampering validation. Hence, the question which model is the best choice for a specific endpoint/exposure route is not easily answered;
(v) Some NAMs are not easy to standardise, in particular those based on omics readout, which generally require complex analysis.

For risk assessment in the food and feed sector, the Scientific Committee of the European Food Safety Authority (EFSA) assessed the state of the art concerning the use of experimental animals and outlined strategies to reduce animal testing (EFSA Scientific Committee, 2014). EFSA also released a concept paper on the use of NAMs with the overall aim to facilitate their incorporation in chemical food and feed risk assessments (EFSA, 2022). A roadmap (Escher et al., 2022) for action to support the implementation of NAMs in risk assessment for the food and feed sector was published, which provides a comprehensive summary of currently available NAMs. The review was organised into five research areas: toxicodynamics, toxicokinetics, exposure, susceptible human population, and data integration. The focus was on the oral route of exposure considering the endpoints of specific target organ toxicity after repeated-dose (sub-chronic to chronic) exposure, endocrine disruption (ED), (non)genotoxic carcinogenicity, (developmental) neurotoxicity, immunotoxicity and reproductive toxicity.

The NAM roadmap did not review the work that has been done for NMs but focused on conventional chemicals. Therefore, within the EFSA NAMS4NANO project, our consortium expanded the existing NAM roadmap (Escher et al., 2022) to provide an overview on existing NAMs that could be suitable for NMs (Usmani et al., 2024). Applying NAMs to NMs poses several additional challenges (e.g., issues with dispersion stability, dosimetry, agglomeration, dissolution and potential assay interferences) but also offers specific advantages beyond the fact that NAMs are generally well suited to keep pace with innovation and the steadily increasing number of new (nano-)materials. The following opportunities for NMs are worthwhile to consider:

- The existing OECD Test Guidelines (TGs) are not suitable as such, requiring adaptations for nano-scale considerations.

- Some essential elements for risk assessment of NMs, such as the identification of nanoparticles (NPs) inside cells and tissues and the assessment of cellular uptake and/or barrier crossing, are technically easier to implement *in vitro* rather than in *in vivo.*

- Nanotechnology is an area with high level of innovation with established dialogues for different stakeholders such as researchers in academic and public institutions and

industry. New concepts such as "safe by design (SbD)"/ "safe and sustainable by design (SSbD)" or NAM-based Integrated Approaches to Testing and Assessment (IATAs) for improved mechanistic understanding have received considerable attention by all stakeholders.

In the food and feed sector, the term "engineered nanomaterials" is defined within the Novel Foods Regulation (EU) 2015/2283 as follows (EC, 2015).

> The term "engineered nanomaterial" is defined in the EU Regulation on Novel Foods (EU) 2015/2283 as "any intentionally produced material that has one or more dimensions of the order of 100 nm or less or that is composed of discrete functional parts, either internally or at the surface, many of which have one or more dimensions of the order of 100 nm or less, including structures, agglomerates or aggregates, which may have a size above the order of 100 nm but retain properties that are characteristic of the nanoscale." Nanoscale refers to a size range between 1 to 100 nm.

However, nano-specific considerations in the risk assessment are not limited to NMs but may also apply to conventional materials, which contain a nanoscale fraction. Therefore, within this document, the term "nano" shall be used for all materials that would require nano-scale considerations in their safety assessments, in line with the EFSA Guidance (EFSA Scientific Committee, 2021a; 2021b). Specific considerations on the use of NAMs for the risk assessment of NMs are described in the EFSA Guidance Nano-RA (EFSA Scientific Committee, 2021a), as summarised in Figure 1. It suggests that nano-specific risk assessment is best achieved through IATAs with NAMs as the first choice to generate new information.

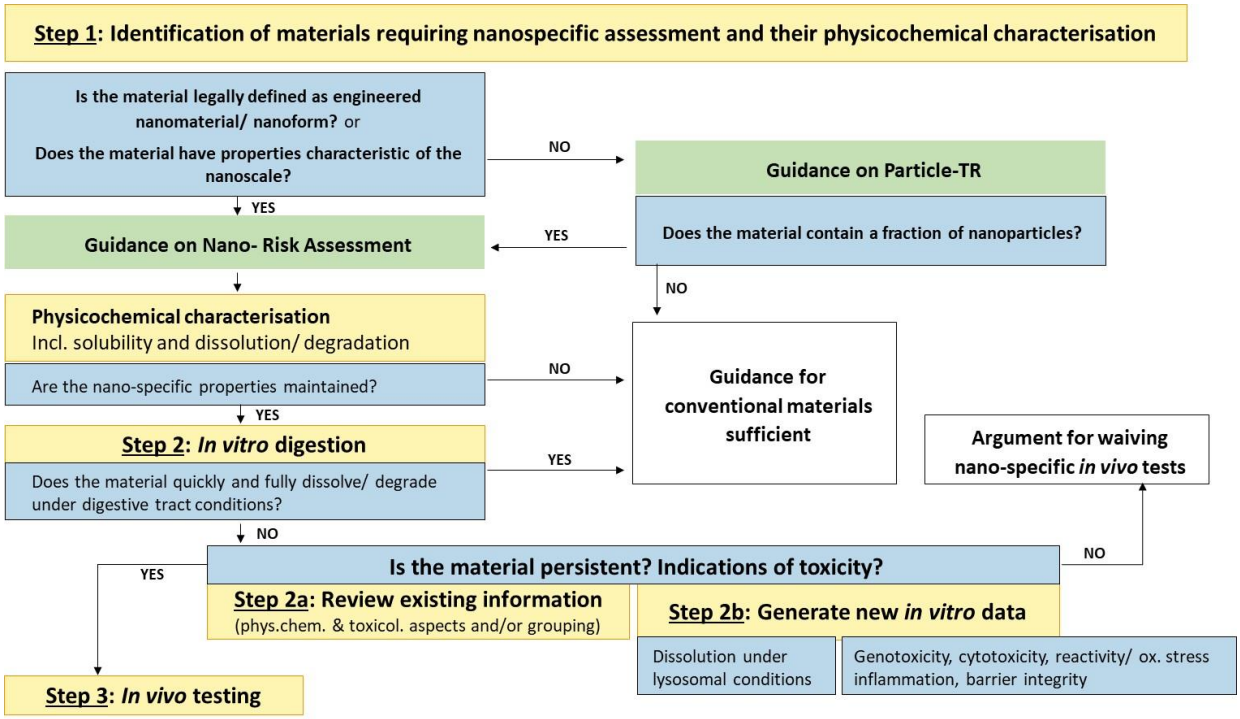


Figure 1: General approach to the risk assessment of NMs as outlined in the EFSA Guidance on Nano-RA (Drawn according to: EFSA Scientific Committee, 2021a)

It begins with physicochemical characterisation to determine whether the material potentially requires nano-specific considerations in its risk assessment. After characterisation, solubility and dissolution/ degradation behaviour of the material in relevant biological fluids is assessed to determine whether the nano-specific characteristics of the test material are maintained after uptake into the body. If this is the case, the next step asks to review all the existing information on toxicity (in vitro and in vivo) along with specific physicochemical characterisation data. Information on carcinogenic, mutagenic, reprotoxic (CMR) properties should always be included. If this information is not sufficient to conduct a risk assessment, in the next step new in vitro studies using a battery of relevant in vitro toxicity tests considering genotoxicity, cytotoxicity, reactivity/ oxidative stress, (pro-)inflammatory responses and intestinal barrier impairment shall be conducted to address gaps and/or uncertainties. In addition, dissolution shall be investigated under lysosomal condition. Importantly, if a material demonstrates absence for persistence based on dissolution data using simulated lysosomal and gastrointestinal conditions and shows no indication of potential toxicity (based on existing data and/or the in vitro test battery), these can serve as arguments for waiving further nano-specific in vivo tests. For cases where new in vivo studies are required, the review of the existing data along with the data obtained in the NAM-based tests will guide the design of these in vivo studies to avoid unnecessary animal testing. Therefore, NAMs are prioritised for generation of de novo information, being tailored to address questions specific to the material being evaluated in risk assessment. IATAs are proposed for the integration of human, animal and NAM data.

Table 1 provides an overview on NAMs identified relevant for NMs in our review (Usmani et al., 2024) addressing Step 2b (refer to Figure 1) according to the EFSA Guidance on Nano-RA (EFSA Scientific Committee, 2021a).

Table 1: Overview of NAMs identified relevant for the first steps of NM risk assessment (Usmani et al., 2024)

| | NAMs (total) | Validated (chemicals) | Under validation (chemicals) | Validated (NMs) | Under validation (NMs) | Under Development |
|---|---|---|---|---|---|---|
| Genotoxicity | 34 | 5 | 3 | 0 | 1 | 25 |
| | - | OECD TGs 471, 473, 476, 487, 490 | WNT Projects 4.139, 4.125 4.139 | - | SPSF accepted (OECD TG 487) | - |
| Cytotoxicity | 40 | 5 | 0 | 6 | 0 | 22 individual 7 multiple |
| | - | OECD TGs 431, 432, 439, 498 OECD GD 129 | - | ISO 19007:2018 ISO/TR 22455:2021 ISO/TR 21624:2020 ISO/TR 19601:2017 ISO/TS 21633:2021 ASTM E2526-22 | - | - |
| Reactivity/ Oxidative Stress | 22 | 1 | 0 | 4 | 0 | 16 individual 1 multiple |
| | - | OECD TG 495 | - | ISO TS 18827:2017 ISO 20814:2019 ISO/TS 19006:2016 ASTM E3351-22 | - | - |
| (pro-) inflammatory Responses/ Barrier Integrity | 12 | 0 | 0 | 0 | 0 | 7 individual 5 multiple |

Under development – Methods established in nanosafety projects, most with SOPs. Individual –method description covers one method only. Multiple –high throughput method, several (i.e. multiple) readouts for different endpoints.

## Developing NAMs for risk assessment

Overall huge efforts have been undertaken to establish a variety of NAMs in the last decades, many of which are meanwhile routinely applied in fundamental and applied research but only a few have been implemented in risk assessment. Developing a NAM-based test method for chemical risk assessment is a time- and resource-intensive process that involves several phases as summarised in Figure 2.

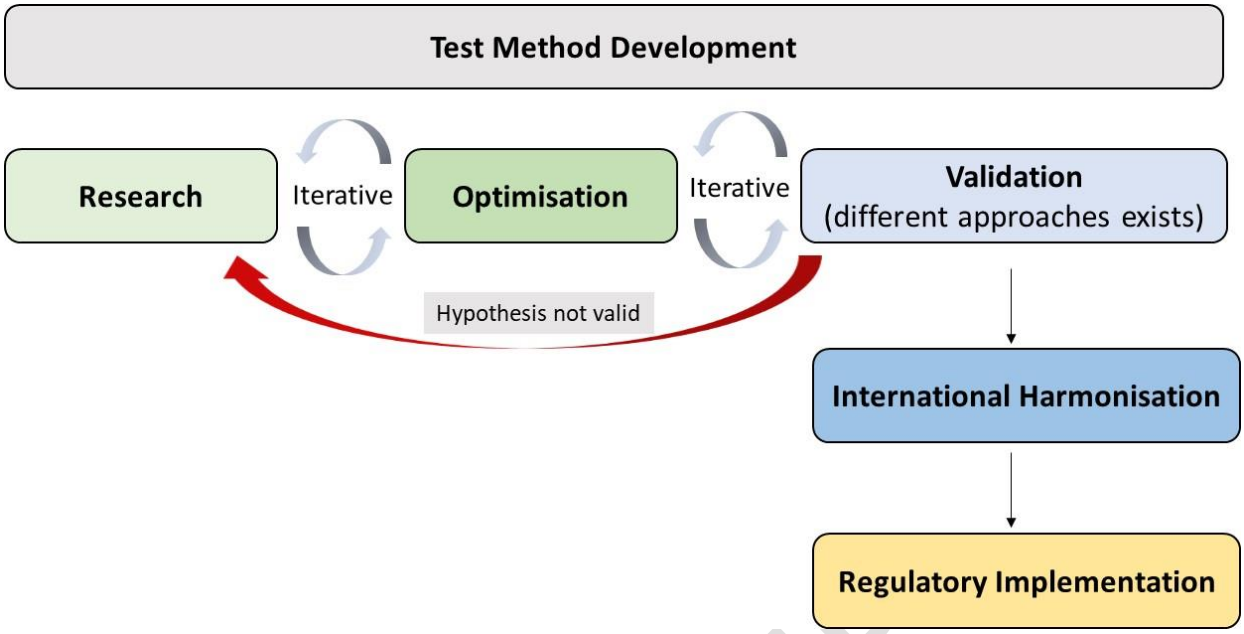


Figure 2: The development of a test method for regulatory purposes typically involves four major phases, namely research, optimisation, validation and international harmonisation before it is "ready" for regulatory implementation

Within this document the term "test method development" is placed in a broad context to cover all the phases until the method is "ready" for regulatory application. Hence, four major phases can be distinguished: i) research, ii) optimisation, iii) validation and iv) international harmonisation. For chemical risk assessment this typically means the endorsement of the method as a TG by the OECD. It should be kept in mind that these phases are not always clearly separated. Often iterative cycles are required and the phases also may (partially) overlap. For instance, methods are often submitted to the OECD before their validation is completed such that validation and international harmonisation are approached together.

The research phase is mainly driven by novelty and often with the motivation to describe the findings in peer-reviewed scientific publications. It ends with a "scientifically plausible" method. Also, a method description is available at this stage along with selected performance parameters. However, before the method can be considered for regulatory use, it needs further optimisation. Most importantly, the method description needs refinement to ensure reproducibility (e.g., to allow for transferring the method to other laboratories without misunderstandings). At this stage, the specific regulatory application (how the method will be applied, for which purpose/ in which context) is often not really clear. The optimisation ends with a "harmonised" or "standardised" method description, often released as SOPs. In the next phase, validation is conducted to test one or several hypotheses formulated on the basis of biological, analytical and *in silico* test results. It is important to note that if the hypothesis is not valid, the test development has to start again from the research phase (Figure 2). Validation demonstrates that the method is "fit-for-purpose" by proofing its reliability and its relevance. Overall, different validation approaches exist as described in detail in the next chapter (1.3.3). Following validation, different routes for regulatory implementation exist. Usually, international harmonisation is regarded very important. The most common route for

regulatory implementation of test methods for chemical risk assessment are OECD TGs, as described in the following chapter (1.3.4). However, also other routes for international harmonisation exist. Examples are the guidelines released by the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) or the standards of the International Standardisation Organisation (ISO).

### Validation of NAM-based test methods

Method validation is a process to provide objective evidence by demonstrating its relevance and reliability to meet the specific requirements associated with its intended use. In the context of validation, the term "method" is broadly understood to cover various types of methods (i.e., *in silico*, *in chemico*, analytical, biological). These methods may be in different stages of method development with some even being commercialised.

All methods have to undergo validation as any analysis or measurement is only worthwhile if the outcome is shown to be accurate for the intended use. Otherwise, the outcome is meaningless. However, the choice of the appropriate validation approach depends on the type of method, its intended use and is critically influenced by other factors such as the available resources. Test methods can be validated by different actors and the validation process can range from "in-house" validation in the research phase to large interlaboratory comparisons (ILCs) conducted for regulatory implementation (typically overseen by a validation body such as the European Reference Laboratories). Therefore, validation is important for all stages of method development, not only for regulatory implementation. One of the important questions to clarify during validation for regulatory implementation is how the results (e.g. from an *in vitro* test procedures) are going to be converted into a prediction that is useful for risk determinations. Often, this question is difficult to answer.

The EU ToxRisk project[2], which aimed to advance the paradigm shift in toxicology towards mechanism-based testing, has identified several challenges that hinder the validation of NAMs (Gourmelon et al., 2024). The project focused on two endpoints, repeated-dose toxicity (RDT) and developmental and reproductive toxicity (DART). Even though for those, the use of NAMs is encouraged in different EU chemicals legislations, the current standard test requirements are *in vivo* studies. The project therefore designed several case studies to explore how and to what extent NAMs can be used for hazard and risk assessment. In this context, the authors assessed method validation for regulatory implementation (Gourmelon et al., 2024) and identified several challenges, *inter alia* validation requires significant resources, it is non-appealing to academic researchers, there is limited knowledge on how to conduct a validation study and finally there is reluctance to invest in validation of NAMs without a defined application (Gourmelon et al., 2024).

In the following, we briefly summarise important aspects for the validation of the various types of NAMs, which are of relevance for our proposed qualification system, namely *in silico*, *in chemico*/ analytical and biological (*in vitro*, *ex vivo* and whole organisms) NAMs.

[2]www.eu-toxrisk.eu

## Validation of *in silico* NAMs

In the last decades, technological development has resulted in different types of *in silico* NAMs being established for chemical risk assessment. The most important are Physiologically Based Kinetic (PBK) and (Quantitative) Structure Activity Relationship ((Q)SAR) models. Although most of them are not yet established or integrated in OECD TGs, OECD guidance documents (GDs) are published which summarise important aspects concerning the validation of *in silico* NAMs.

PBK models simulate the uptake and distribution of chemicals within body compartments with the aim to characterise their toxicokinetic properties. In NAM-based testing approaches, PBK models are essential for *in vitro* to *in vivo* extrapolations, which converts the effective *in vitro* concentrations into *in vivo* equivalent doses. PBK models are increasingly used for chemicals but they are only emerging for NMs. As elaborated in our review (Usmani et al., 2024), the currently existing PBK models for NMs cannot yet be considered mature enough for implementation in risk assessment. However, as PBK modeling approaches evolve, it becomes crucial to standardise them. Hence, several guidance documents are available that address the characterisation, validation and reporting of PBK models (e.g. EFSA, 2014; OECD, 2021b). In general, any PBK model should be only as complex as necessary. In some cases, simple (one-compartment) models might be sufficient (e.g. for describing uptake across a barrier) while in other cases, more complex (multi-compartment) models are required (e.g. to estimate a systemic concentration from a given exposure). The distribution of a chemical may be either perfusion-limited (with the blood flow rates as the limiting process) or permeability-limited (with permeability across the cell membrane as the limiting process). In some cases, additional specific transport processes are involved and have to be included in a permeability-limited model. Any PBK model contains two types of parameters, anatomical and physiological parameters for the species of interest (such as tissue volumes, blood flow rates) and chemical-specific parameters related to absorption, distribution, metabolism and excretion (ADME) such as partition coefficients, rates of absorption, binding constants, elimination rates. The latter can be obtained in different ways (from literature, experimentally (*in vivo*/*in vitro*) or by modelling/ estimation). The OECD PBK guidance (OECD, 2021b) elaborates on different approaches for determining the required parameters (e.g. the absorption parameters) and mentions critical issues that should be considered when setting up, using or assessing a model. A typical critical issue is the scaling of *in vitro* ADME properties to the *in vivo* situation. The use of different *in vitro* models for the individual properties can also affect the modeling results. An example is intrinsic hepatic clearance, which might be measured using different NAMs employing different cell models with a different metabolic capacity. Importantly, for several of the chemical-specific parameters, nano-specific considerations are required (e.g. NMs may have different uptake mechanisms, or some simple mono-culture cell models might be of limited use). We have elaborated on these and other nano-specific issues in our review (Usmani et al., 2024). However, in the existing guidance documents, nano-specific considerations are not yet included.

QSAR methods describe mathematical relationships, which can be applied to predict missing physicochemical properties and toxicity endpoints. They are already frequently applied for screening and prioritisation. In order to enhance the regulatory use of QSAR methods in

chemical risk assessment, the OECD has established the QSAR toolbox[3]. Furthermore, it has released a QSAR assessment framework for evaluating the scientific validity of (Q)SAR models, predictions, and results (OECD, 2023c). It distinguishes between the model itself (i.e. setting up the model), the application of the model and the results. The model developer should report all information about the model(s) using the (Q)SAR Model Reporting Format document. The users of the model should report all the information about the (Q)SAR prediction(s) in the (Q)SAR Prediction Reporting Format document. This then enables the regulator to assess the model(s) and the prediction(s) using the QSAR assessment framework. Concerning the model, the framework emphasises the following aspects: i) a well-defined endpoint, ii) an unambiguous applicability domain, iii) appropriate measures for goodness-of-fit, robustness, predictivity of the model and iv) a mechanistic interpretation. Concerning the application of the model (predictions using the model), the framework identified the following important aspects: verification that correct input parameters are used, that the substance is within the applicability domain of the model and the reliability (being linked e.g. to reproducibility as it could be assessed by applying the same model to similar substances or e.g. by mechanistic considerations). The QSAR assessment framework does not yet include any nano-specific considerations. It should be noted that currently, the OECD QSAR toolbox does not contain any nano-QSAR models. However, all the parameters mentioned above are equally true for nano-QSAR models.

### Validation of *in chemico* and analytical NAMs

Analytical and *in chemico* NAMs are experimental approaches, which do not require a biological model. Examples for *in chemico* NAMs are dissolution/degradation rate tests (e.g. dissolution in water, in GIT fluids, in lysosomal conditions), reactivity assays (e.g. NM surface reactivity), and binding assays (e.g. receptor-ligand). Several OECD TGs are established for *in chemico* NAMs specifically developed for NMs. An example of a recently published one is the TG on the determination of the Hydrophobicity Index of NMs through an affinity measurement (OECD 2023d).

Analytical NAMs describe analytical measurement procedures such as those commonly applied for NM physicochemical characterisation or those utilised to quantify migration/release. They may be seen as part of the wider group of *in chemico* NAMs. For analytical NAMs, organisations such as the International Organization for Standardization (ISO) and the European Committee for Standardization (CEN) operate in the field of establishing analytical standard methods. A few of the ISO and CEN standards are relevant for risk assessment and hence have been implemented for this purpose in different legal frameworks. In the EU, in the food and feed sector, analytical methods are also established by European Reference Laboratories (EURL) and, in the area of NMs, by a specific initiative of the EC JRC, supported by a group of expert labs (Nanomaterials in Food Laboratory Group/ NIF-LAG). However, it is evident that a standard method cannot be established for every analyte and/or every matrix. Therefore, in-house analytical methods validated according to available guidance are routinely applied for risk assessment in the food and feed sector.

[3]https://www.oecd.org/chemicalsafety/risk-assessment/oecd-qsar-toolbox.htm

All these validation approaches aim to demonstrate that the analytical method is suitable for its intended purpose, to characterise its performance characteristics (e.g. specificity/ selectivity, working range, accuracy and precision, robustness, LoDs/LoQs), and determine the measurement uncertainties. The 'Guide to the expression of uncertainty in measurement' (GUM) establishes general rules for evaluating and expressing uncertainty in measurement (ISO/IEC Guide 98-1:2024). These principles are intended to be applicable to a broad spectrum of measurements.

Prerequisite is a proper method description (preferably via SOPs) with sufficient level of details (e.g., how the samples are to be prepared, what is measured, how is it measured, how is the data analysed). The scope of the method needs to be given. This includes which concentration range is adequate or which matrices can be analysed. Each validation requires a detailed validation plan. In general, during the validation of an analytical method the following parameters have to be determined (Linsinger et al., 2013):

**Linearity** – the ability of a method to obtain test results proportional to the value of the measurement, which needs to be demonstrated by a calibration over the intended working range of the method.

**Working Range** – which is defined by providing lower & upper limits, between which the analytical procedure provides an acceptable linearity, trueness and precision.

**Limit of Detection (LoD)**/ **Limit of Quantification (LoQ)** - this refers to the lowest concentration or size that can be reliably detected (LoD)/ reliably quantified (LoQ).

**Trueness** - is assessed by estimating the magnitude of the experimental bias and by determining if it is statistically significant. Bias can be determined by using certified reference materials, by comparing the new method with a reference method or by making use of data coming from an ILC.

**Selectivity** – the ability of a method to correctly determine the analyte without interference from other compounds.

**Precision** – is defined as the closeness of agreement among individual test results from repeated analyses of a homogeneous sample. It can be determined at three levels: repeatability, intermediate precision and reproducibility.

**Repeatabilty** – refers to the ability of the method to generate the same results over a short time interval under identical conditions. It is recommended to measure at least ten subsamples in one run or three subsamples over five different days.

**(Intermediate) precision** – refers to the agreement between the results from within-laboratory variations due to random events that might occur when using the method such as testing on different days, involving different analysts, using different equipment and/or under different calibration. At least the variation of one of these factors has to be investigated. These studies are carried out according to ISO 5725-3 (2023).

**Reproductibility** - refers to the results of collaborative studies among different laboratories (ILCs). Documentation in support of reproducibility studies should include the standard

deviation, the relative standard deviation (or coefficient of variation), and the confidence interval.

**Robustness** - is defined as a measure of its capacity to obtain comparable and acceptable results when perturbed by small but deliberate variations in procedural parameters listed in the documentation. During a robustness test small but deliberate changes are made to the method variables (e.g. temperature) to study the effect on the method performance.

The EU NanoDefine project worked on a guidance for the in-house validation of nano-specific analytical methods with the goal to prepare the methods to become standard methods. The three parts of the resulting NanoDefine Methods Manual aim at providing guidance through the NM characterisation process, on the use of the characterisation methods, their application range and their limits to assist the user in choosing the most appropriate measurement method(s) to identify whether or not a material is a NM according to the EC recommendation for NM definition (2022/C229/01). The focus lies on characterisation methods, which are candidates for performing a reliable analysis of the number-based size distribution of a particulate material, with the goal to identify NMs. Within NanoDefine, several methods have been validated using this approach. Examples include the evaluation and validation of a TEM based approach for size measurement of particulate (nano)materials (Verleysen et al., 2019) and the validation of a particle tracking analysis method for the size determination of nano- and microparticles (Kestens et al., 2017).

### Validation of biological NAMs

The validation of biological NAMs is more complex than that of *in chemico*, analytical and *in silico* NAMs as an unlimited number of combinations are possible when considering the huge variety of biological models (e.g., established cell lines, primary cells, co-cultures, organoids, *ex vivo* structures, whole organisms) each of which can be investigated using different measurement procedures (e.g., by spectrophotometric measurements using simple plate readers, employing various fluorescence measurements in flow cytometry or by advanced imaging (including high content methods that multiplex different readouts) or omics methods). Moreover, biological NAMs can measure different biological events, which may appear at different levels of biological organisation including the molecular level (e.g., a ligand binding to a specific receptor), cellular (e.g., the activation of a specific cell population) or considering events that occur on different tissue-levels (e.g., exposure to a sensitising chemical on the skin leading to T-cell stimulation in the lymph node). All of this renders the validation of biological NAMs much more challenging and complex compared to analytical and *in silico* NAMs.

The European Reference Laboratory for alternatives to animal testing (EURL-ECVAM) has proposed a modular approach for validating biological NAMs (Hartung et al., 2004, Figure 3). In total, seven modules are described for validation. It starts with the test definition, which includes a detailed description of all procedures, preferably in SOPs. Next, several modules address "reliability", namely the assessment of the within-laboratory reproducibility of a method, the transferability of the test to other, method-naïve laboratories and finally the between-laboratory reproducibility, which is typically assessed in blind interlaboratory comparison (ILC) studies. The next two modules address "relevance". In this context the

predictive capacity and the applicability domain of the method have to be described. The last module (performance standards) is optional but useful. It suggests defining performance standards to facilitate the validation of similar methods in future. Importantly, the validation results shall be assessed by an independent expert panel.

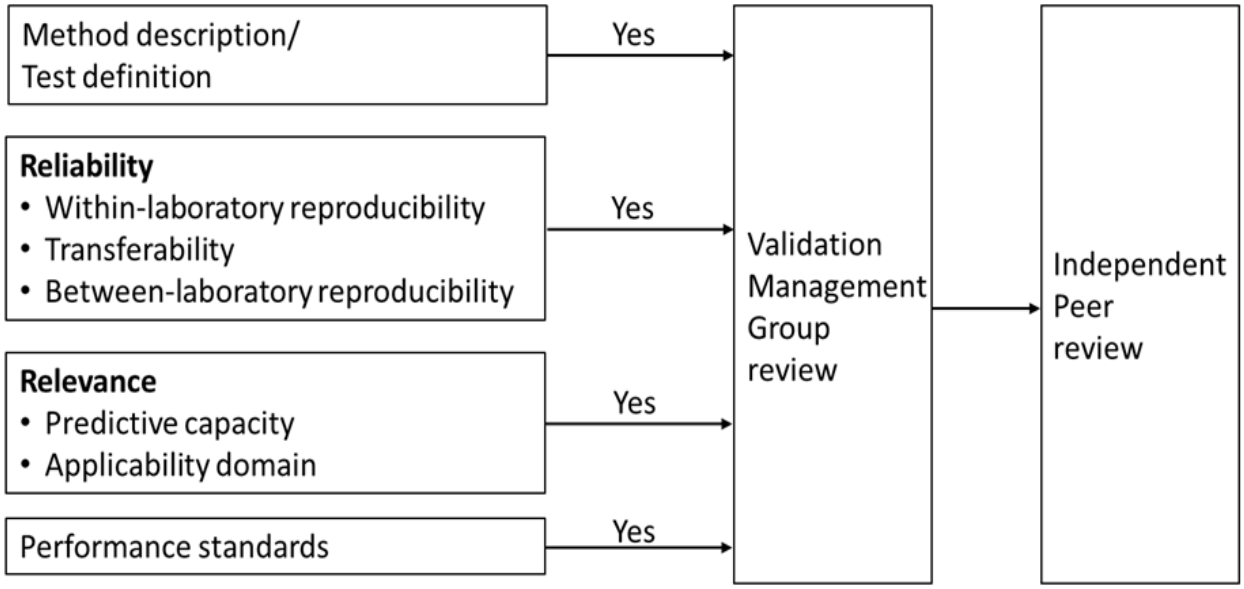


Figure 3: The EURL-ECVAM modular validation approach (adapted with modifications from Hartung et al., 2004)

Also, the EURL-ECVAM test submission template[4] provides guidance on how to structure information in order to facilitate validation. Specific aspects covered in the template include information on mechanistic relevance of the method, biological relevance of the test system, intended purpose (stand-alone or within integrated and/or defined approach), limitations of the test method, controls (positive, negative, and benchmark), quality acceptance criteria, reproducibility and predictive capacity of the method. Further guidance on the validation of non-animal methods is included in the EURL-ECVAM status report (JRC, 2023). For instance, it provides resources for developers on validation and qualification of Organ-on-Chip devices that were developed in collaboration with European Organ-on-Chip Society[5]. The resources include the EFSA guidance on assessment of peer reviewed open literature results (EFSA, 2010). In addition, frequently asked questions on regulatory acceptance of Organ-on-Chip are compiled and structured according to different domains such as improvements needed to establish alternative methods.

Further practical aspects concerning the validation of toxicity test procedures were compiled in an ECVAM workshop report (Balls et al., 1995). The authors suggested classifying validation of toxicity test procedures into three categories depending on the intended use of the method. The first category of validation is for non-regulatory studies that are applied in screening. These can include methods that provide some mechanistic information relevant to adverse

[4]https://joint-research-centre.ec.europa.eu/document/download/34e99b98-cd37-40d2-9779-382921ca04b6_en?filename=TPF%20Test%20Presubmission%20Form.docx
[5]https://data.jrc.ec.europa.eu/dataset/7bcb1db5-5c7e-460b-b79e-ca5f642514a4#details

effects but do not have to provide a full coverage of all potential mechanisms leading to e.g. systemic toxicity. The second type of validation is for methods that are intended to be included in regulatory risk assessments but in hierarchical approaches. The third type is for methods which should replace an existing animal study. In this case, there needs to be sufficient confidence that the method provides equivalent or better level of protection than the existing approach. Balls et al. (1995) mention a possibility of a fourth type wherein *in vitro* method provides part of data required for the regulatory endpoint which will eventually "accrue" sufficient data to support replacement of animal test.

## Regulatory Implementation: OECD TGs

The most universal and common approach towards regulatory implementation of test methods for chemical risk assessment is to achieve endorsement as OECD TGs. These TGs are the most universally applied in the context of chemical risk assessment. They are specifically developed for international use and are considered appropriate for safety testing of chemicals. According to the OECD Council Decision on the Mutual Acceptance of Data (MAD) of 1981, that was recently amended (OECD, 2024), the data, which are generated in accordance with OECD TGs and the OECD Principles of Good Laboratory Practice (GLP, 1998) will be accepted by regulatory agencies in other OECD member or adherent countries for purposes of chemical risk assessment.

The OECD TGs are released within the Environment, Health and Safety Programme and are categorised in five sections, namely i) physicochemical properties (>25 TGs), ii) biotic systems (>50 TGs), iii) environmental fate (>20 TGs), iv) health effects (>80 TGs) and v) residual chemistries and others (9 TGs). In the following, we consider only non-animal-based TGs. **Classical OECD TGs** include a single method only, which has been validated using an appropriate number of chemicals. **Performance-based OECD TGs** describe at least one formally validated reference method and in addition, performance standards are included. This means, the essential test method components are identified, a list of chemicals to demonstrate comparability of results is provided and minimal performance criteria are established. Thereby, additional methods can be included later on in the same TG, which are then validated only according to the performance standards. Importantly, all the methods in such TGs measure the same target/endpoint and are technologically and functionally similar. Examples for performance-based OECD TGs are OECD TG 431 (2019a) on skin corrosion and OECD TG 492 (2023e) on eye damage. **Key event (KE)-based TGs** also contain several validated methods. They address the same KE but are not technologically and functionally similar. The OECD TGs 442C (2023f) and 442D (2022) for skin sensitization may serve as examples for KE-based TGs. Finally, **Defined Approach (DA) TGs** are similar to KE-based TGs as they include several validated methods that are not technologically or functionally similar. In contrast to the KE-based TG, the methods in DA TGs are used in a specific combination to predict an adverse effect, and furthermore the TG defines fixed interpretation procedures. The OECD TG 497 (2023g) is an example of DA for skin sensitization.

In order to establish an OECD TG, a standard project submission form (SPSF) is submitted to provide the regulatory need for the method along with information such as the relevance for more than one country, the applicability domain (i.e., for which classes of chemicals is the method suitable) and whether this method shall be a new TG or an update for an existing TG.

Most importantly, the validation status of the proposed method needs to be given including an action plan and specific timelines for each remaining step. Ideally, a method is fully validated by a validation center such as EURL-ECVAM, ICCVAM, JACVAM etc. before submitting a SPSF. However, in reality often SPSFs are submitted for methods, which have not yet been validated. In such a scenario, validation studies could be organised at the level of the OECD. However, this renders the process more time-intensive.

OECD TGs are living documents, meaning they can be revised to keep pace with scientific developments or deleted if obsolete. Revised or deleted TGs should no longer be used for testing but OECD provides an overview of them as they are referenced in old study reports.

The principles for validation and international acceptance of new or updated test methods for hazard assessment are described in OECD Guidance Document (GD) 34 (OECD, 2005). Recent projects that dealt with innovative methods revealed that new ways might be needed to conduct validation. These issues have been discussed in several workshops (OECD, 2023h) and will be addressed in the currently ongoing revision of OECD GD 34. Important points discussed are i) how to evaluate test method readiness before methods enter validation (as different approaches have been suggested), ii) how to evolve the concept of performance standards (in particular for new endpoints, what is essential to be determined and what could be optional and included later as the project is progressing), iii) whether to develop a dedicated section of TGs that are not stand-alone methods, iv) how to develop guidance for validation of test batteries and their building blocks and/or validation of new technologies (e.g., Organ-on-Chip), v) how to optimize the recruiting of laboratories for validation studies and vi) how to better report study results.

## Regulatory Use of Non-Guideline Methods

A huge number of promising NAMs are currently available, but only a small fraction has been established as OECD TGs for implementation in a regulatory context. Hence, many NAMs are widely established and applied for use in risk assessment and regulators are increasingly faced with data from non-guideline studies. Though non-standard methods can be used for risk assessment, they require complex and time-consuming case-by-case evaluation. In an ideal case, these non-guideline NAMs are established and described as suggested in the various guidance documents. For *in silico* NAMs, these requirements are summarised in the QSAR assessment framework (OECD, 2023c) and the PBK guidance (OECD, 2021b). Analytical (*in chemico*) NAMs should be at least in-house validated following the principles described in the ISO GUM (ISO/IEC Guide 98-1:2024). Biological NAMs should be described following the OECD GD 211 (OECD, 2014).

OECD GD 211 is the key resource for regulators and method developers for applying non-guideline cell-based methods in risk assessment (OECD, 2014). It lists main issues sorted into five categories, i) general information (e.g., assay name, summary, key reference, status of development), ii) test method definition (e.g., purpose, scientific principle, tissue/ cell, exposure regime, response measurement, quality/ acceptance criteria), iii) data interpretation, iv) test method performance (e.g., robustness, reference chemicals, performance measures, scope and limitations) and v) potential regulatory applications.

Within the EU Project ToxRisk[6], a survey of the scientific literature revealed that method descriptions still show an enormous heterogeneity in quality, detail and scope (Krebs et al., 2019). Two main issues were identified with respect to compliance with OECD GD 211. Firstly, the GD 211 is a highly condensed document, in which single questions often relate to distinct features, which may not be obvious without further guidance. It was suggested to subdivide the questions such that each single question addresses only one unique aspect. Furthermore, it was recommended that the questions should be better explained with additional guidance for each of them. Secondly, the reporting of the method was not standardised. Thus, even for methods which in general comply with GD 211, it was difficult to find the relevant piece of information for each aspect in the publication. Therefore, it was suggested to provide a template for reporting non-guideline methods. Both aspects were addressed by ToxTemp, which is a template for the description of cell-based toxicological test methods based on GD 211 to allow for evaluation and regulatory use of data from non-guideline methods (Krebs et al., 2019).

Another important overarching OECD GD for cell-based toxicological tests is the Guidance on Good *In Vitro* Methods Practice (GIVIMP), released in 2018 (OECD, 2018). It provides guidance for the development and implementation of *in vitro* methods for regulatory use. It was developed as a reference for best practices and as a tool to avoid a reproducibility crisis in *in vitro* toxicological science. It covers several important aspects including i) general quality considerations, ii) requirements for the laboratories, equipment, materials and reagents with recommendations how to improve the reproducibility, iii) practical details for different biological models along with best practices for storage, handling, authentication and characterisation of cell and tissue-based test systems, iv) the importance of reference/control items and how to best assure accurate and reliable exposure to avoid various interferences, vi) the importance of SOPs to simplify the work and to reduce variability due to deviations from a fixed methodology, vii) considerations on method performance and viii) guidance for reporting the results.

## Enhancing Regulatory Use of Non-Guideline Methods: Qualification

Qualification systems have been established to speed-up the regulatory use of non-guideline methods and to enhance confidence in their application for risk assessments in specific sectors and for specific contexts-of-use. Existing qualification systems operate in the context of Drug Development Tools by the Food and Drug Administration (FDA) and research and development into pharmaceuticals by the European Medicines Agency (EMA). In addition, there is a qualification system for medical device development tools (MDDT) established by FDA.

The US FDA Toxicology Working Group has released a Predictive Toxicology Roadmap (FDA, 2017) to describe their current thoughts on how to foster the development and evaluation of emerging toxicological methods and new technologies and how to best incorporate them into FDA regulatory review. This document outlines the concept of method qualification as follows: **within the stated context-of-use, qualification is a conclusion that the results of an assessment using the model or assay can be relied on to have a specific interpretation and application in product development and regulatory decision-**

[6]www.eu-toxrisk.eu

**making.** It is explained that formal validation approaches involve lengthy and expensive processes that may not be necessary for all uses of a particular test. Thus, this document also explains the "context-of-use" concept. The context-of-use refers to a clearly articulated description delineating the manner and purpose of use for the tool (when and how it will be used). Adequately specifying the context-of-use is often a difficult first step towards qualification and regulatory acceptance of new methodologies. Qualification also identifies the boundaries of the available data that adequately justify the use of the tool. Models and assays should be suited for a purpose, and in that context, they will have different applicability, assumptions, and limitations. It is emphasised in the roadmap that high quality data are critical for FDA's ability to reach sound regulatory decisions and to retain the public's trust. Hence, FDA must be able to evaluate the applicability, limitations, relevance, reliability, reproducibility and sensitivity of a test or series of tests to confirm that they are appropriately validated or qualified.

Is a qualification of a method an alternative to its validation? And if it is, under which circumstances? Qualification means to assess whether the method is suitable (i.e. reliable and relevant) for a clearly defined intended purpose with fewer stringent requirements compared to formal validation. Typically, qualification is faster and cheaper compared to validation that aims for establishing an international test guideline. However, it will not be a direct alternative to validation (except in a few cases). Qualification is regarded as appropriate in the early stages of a development process, e.g., for research phase and pre-clinical developments. However, once clinical trials should be initiated it is mandatory to use international accepted test guidelines.

Currently, there are three device development tools (DDT) qualification programs at FDA that can qualify tools of three categories namely on biomarkers, clinical outcome assessment and animal models (FDA, 2020; FDA, 2023). Other DDT qualification programs may be established in future, depending on scientific need and availability of resources.

The EMA[7] offers scientific advice to support the qualification of innovative development methods for a specific intended use in the context of research and development into pharmaceuticals (EMA, 2008). The advice is given by EMA's Committee for Medicinal Products for Human Use (CHMP) based on recommendations by the Scientific Advice Working Party (SAWP). The qualification process leads to a CHMP Qualification Opinion or CHMP Qualification Advice (EMA, 2008).

A CHMP Qualification Opinion on the acceptability of a method for a specific use in the context of research and development is based on the assessment of data submitted to EMA. The evaluation is opened for public consultation from the scientific community before its final adoption. EMA publishes all CHMP qualification opinions.

A CHMP Qualification Advice is given on protocols and methods, which do not fully meet the necessary specifications with the aim of moving them into qualification. The advice is based on the evaluation of the scientific rationale and the preliminary data submitted to EMA. Though EMA does not publish CHMP qualification advice, it may include a letter of support that can be published with the agreement from the sponsor(s) of the novel methodology. The

[7]https://www.ema.europa.eu/en/qualification-novel-methodologies-medicine-development

letter of support is a high-level summary of the novel methodology, context-of-use, available data and ongoing and future investigations.

## Evaluating Method Readiness

The evaluation of "method readiness" might be helpful at different stages of test method development. Several approaches have been proposed, most of which cover only biological NAMs and are established to support the establishment of OECD TGs (i.e., to evaluate if a method is "ready" for validation and/or for submitting a SPSF).

## Evaluating Method Readiness for developmental neurotoxicity (DNT)

The first comprehensive framework to assess method readiness was developed for developmental neurotoxicity (DNT). An urgent need was identified to establish an alternative testing strategy and the methodology was discussed in several workshops. The first workshop explored the potential of applying the adverse outcome pathway (AOP) framework to develop a DNT testing strategy (Bal-Price et al., 2015). A follow-up workshop focused on practical aspects such as criteria to evaluate method readiness (Bal-Price et al., 2018). Clearly defined mid-term objectives were to establish a standardised evaluation system for assay readiness, to define a list of suitable test methods (taking into account these criteria) and to construct an IATA for initial screening and prioritisation for DNT. An important finding from the workshop was that "readiness" can be evaluated from different perspectives, i.e., by academic researchers, test method developers or regulators. The authors proposed readiness criteria allocated into 13 clusters, namely test system, exposure scheme, documentation/SOP, main endpoint(s), cytotoxicity, test method controls, data evaluation, testing strategy, robustness, test benchmarks, prediction model, applicability domain(s) and screening hits. A scoring was proposed for each of the criteria in these categories (Figure 4). Importantly, Bal-Price, et al. (2018) suggested evaluation of a method readiness in two phases. Phase I considers basic features of a test method as they would be covered by academic researchers, which are usually the initial developers of a test method. Phase II assesses aspects considered important for industrial or regulatory implementation of the method.

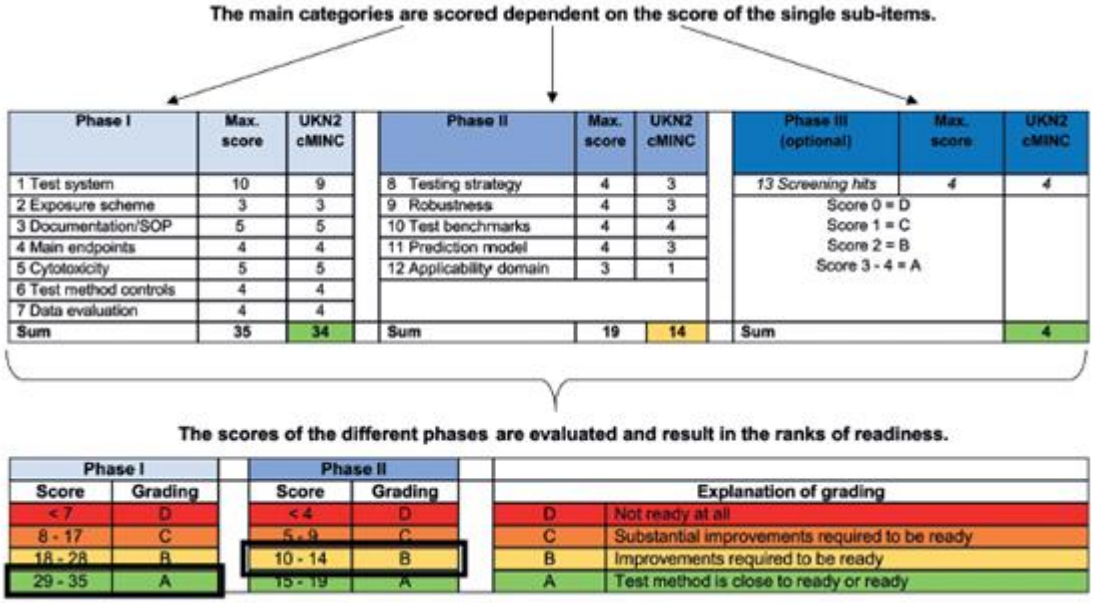


Figure 4: Proposed Scoring System to evaluate method readiness (source: Bal-Price et al., 2018 under open access CC BY creative commons license)

## Evaluating Method Readiness for endocrine disruption (ED)

Several chemical substances have been linked to endocrine disruption (ED), meaning that they can mimic or interfere with hormones resulting in a specific mode of action and ultimately to different adverse effects (e.g., reproductive or neurodevelopmental toxicity, metabolic diseases). Such chemicals are referred to as endocrine disruptors. The current OECD conceptual framework considers chemicals acting on oestrogen, androgen, thyroid and steroidogenic pathways. A few OECD TGs covering ED have been released but overall, there is an urgent need for innovative test methods that can effectively identify and characterise ED. It was recognised that the validation process can be accelerated when essential information requirements on a method were met before the method was submitted to the OECD. Supported by several EU Projects, the French Pepper platform (Figure 5) was established with the aim to accelerate the validation process of methods regarding ED identification[8].



[8]https://eurion-cluster.eu/pepper-call-for-test-methods-submissions-for-2023-2024/

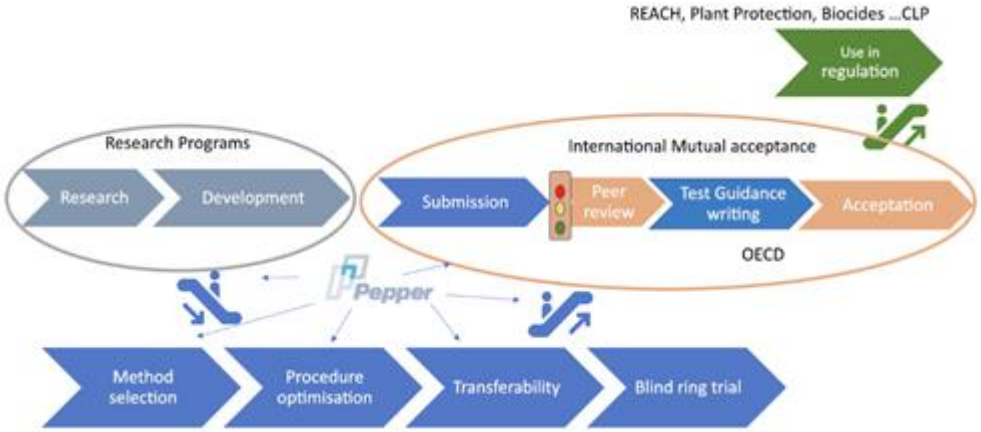


Figure 5: From research to regulation: the role of the Pepper platform in the validation of assays for ED (source: Crouzet et al., 2023 under open access CC BY NC ND creative commons license)

In this context, the ReadEDTest[9] online tool, was developed as a self-assessment questionnaire to evaluate the method readiness using a battery of criteria (Crouzet et al., 2023). The criteria of the ReadEDTest were established according to Bal-Price et al. (2018), ToxTemp (Krebs et al., 2019) and the OECD GD 211. The ReadEDTest covers the following sections: test method description, operational readiness and data management, reproducibility and transferability, historical data, SOPs, relevance of the method and laboratory resources for pre-validation, which are underpinned by sub-sections. For instance, the test method description includes the following sub-sections: test system (i.e., type of model), model characterisation (e.g., metabolic competence), cellular environment (e.g., culture conditions), exposure schemes (for tested compounds) and endpoints (e.g., parameter to be measured).

All sections which are mandatory for validation are scored and a score limit was defined, which needs to be reached in order to enter validation. In contrast, non-mandatory aspects have no score and score limit. In the end, the tool depicts the score to help with the evaluation of the strengths and weaknesses of the method. For instance, a result may conclude that "based on this score the method is not ready to enter validation regarding [the name of the section/subsection]".

Also, the outcome of ReadEDTest can be visualised using radar charts (Figure 6) to identify which sub-sections are ready for validation and which require further development.

[9]https://readedtest.u-paris-sciences.fr/

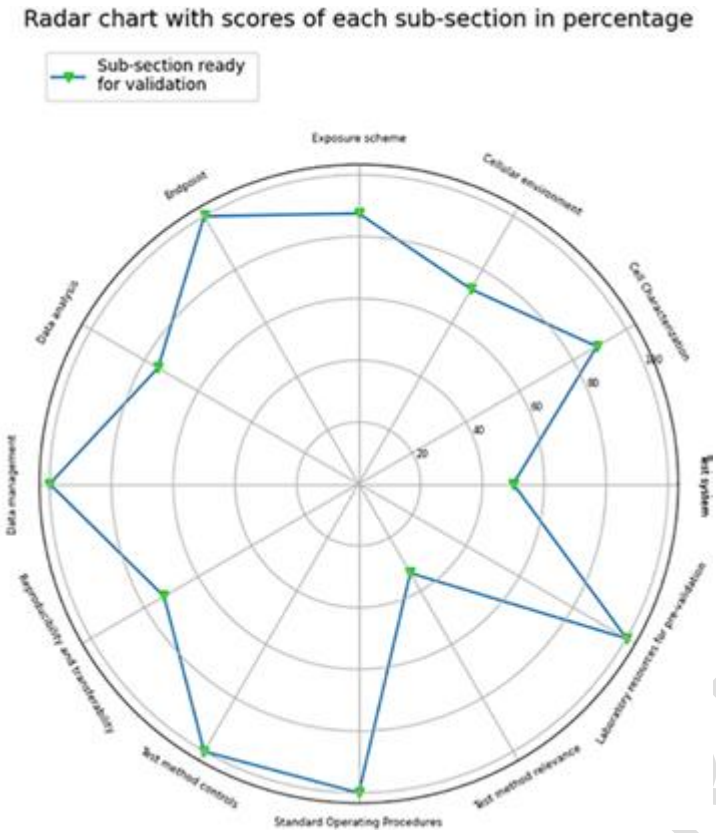


Figure 6: An example of a method (ERα CALUX transactivation assay) evaluated by ReadEDTest (source: Crouzet et al., 2023 under open access CC BY NC ND creative commons license) depicted as a radar chart with green dots indicating sub-sections (cellular environment, cell characterisation, test system, laboratory resources for pre-validation, test method relevance, SOPs, test method controls, reproducibility and transferability, data management, data analysis, endpoint and exposure scheme) being ready for validation while red dots would indicate sub-sections that are not yet ready (here none)

## Transparency, Reliability, Accessibility, Applicability and Completeness (TRAAC) Framework for nano-specific methods

In EU, significant resources have been invested in the development of tools and methods to identify and characterise risks associated with NMs. However, stakeholders such as regulators, policy makers and manufacturing companies may have only limited awareness of these tools and methods. To overcome the lack of regulatory acceptance and the related limited application of such methods, a framework was developed to quantify the regulatory readiness based on the principles of Transparency, Reliability, Accessibility, Applicability and Completeness (TRAAC framework) (Shandilya et al., 2023). For each of the five pillars of the TRAAC framework, clear and quantifiable criteria were proposed to evaluate the overall quality of the methods. The final aim is to give clear quantifiable comparison for any method developed to enable faster regulatory acceptance by increasing trust based on objective data. TRAAC addresses the needs of different communities as it 1) should help regulators to identify and choose the most appropriate methods, 2) assist industries in the application of such

methods and 3) assist method developers to identify gaps in their method and thereby in the development of more fit-for-purpose methods.

The five pillars of TRAAC are:

(i) Transparency: ownership, clear communication about development, methods, strengths and limitations (e.g., boundary of use);

(ii) Reliability: quality, correctness, and consistency of output;

(iii) Accessibility: usability, findability, and user experience evaluation;

(iv) Applicability: applicability domain and adequacy to address user needs;

(v) Completeness: comprehensiveness regarding EU regulatory frameworks (Regulation on the registration, evaluation, authorisation and restriction of chemicals (REACH)) and their requirements for NMs.

For each of the five pillars a set of dedicated criteria has been described (Figure 7).

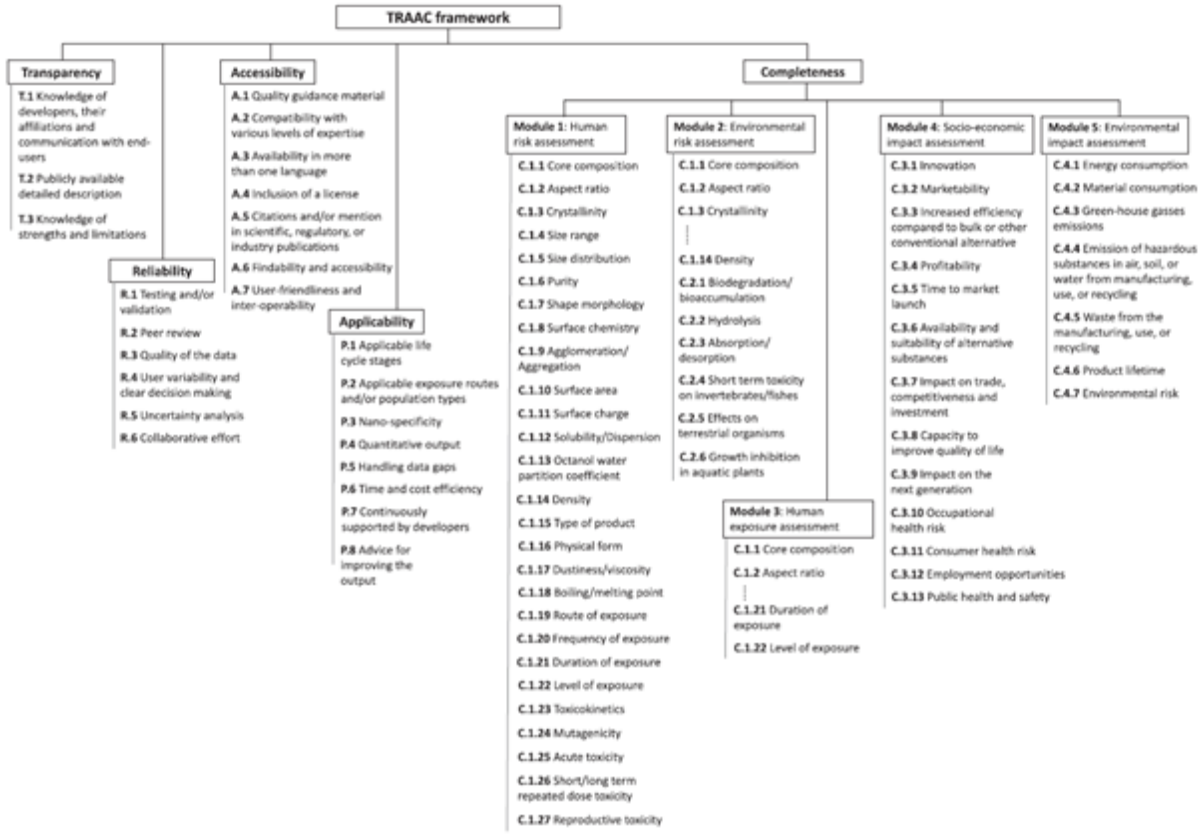


Figure 7: Proposed TRAAC scheme to evaluate method readiness (source: Shandilya et al., 2023 under open access CC BY NC ND creative commons license)

In addition, for each of the five pillars and the different criteria an evaluation system has been formulated which relates to scores (0.1, 0.5 or 1) and weights (1 or 5) leading to a description of the regulatory readiness of the method (Figure 8).

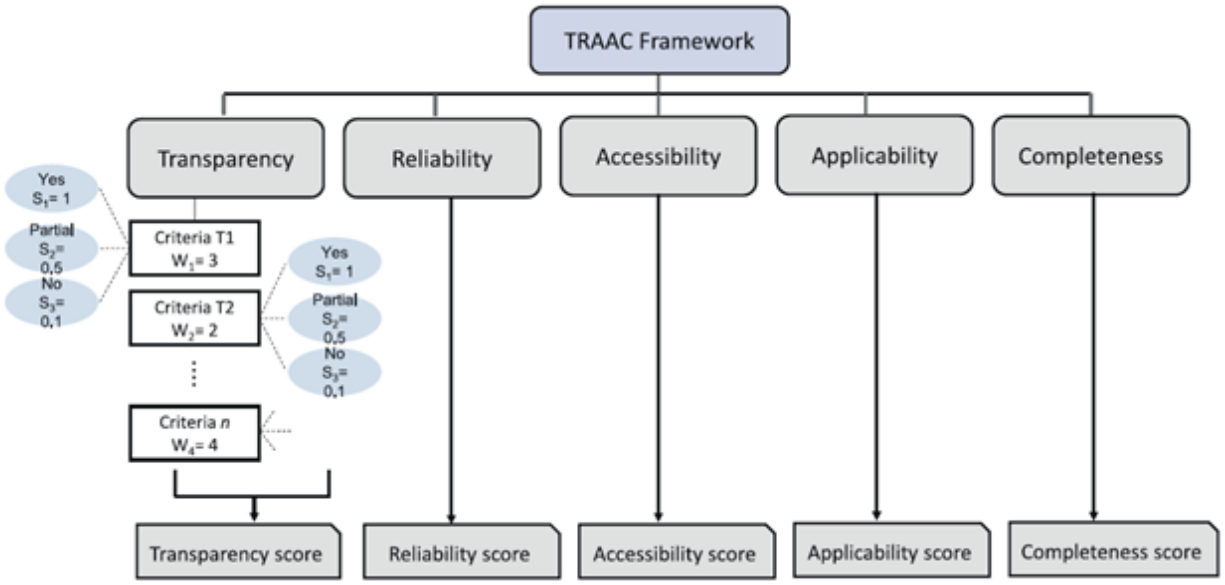


Fig. 2. Scoring system in TRAAC framework.

Figure 8: TRAAC evaluation and scoring scheme (source: Shandilya et al., 2023 under open access CC BY NC ND creative commons license)

The TRAAC system was developed based on the outcome of several EU funded nanosafety projects (especially EU Gov4Nano[10]) considering available regulatory frameworks, stakeholder needs and methods to assess risks for humans and the environment from NMs. The TRAAC system was tested in the EU DIAGONAL[11] project for its applicability.

## Aim of this work

According to the call text (GP/EFSA/MESE/2022/01)[12], the NAMS4NANO project is tasked to develop a proposal for a qualification system that can facilitate the regulatory use of NAMs in the food and feed sectors using the experience of existing qualification systems (i.e., EMA and US FDA). The proposed system should allow submitters of methods/ tools to present proposals to EFSA, supported by scientific evidence on the capacity of the method/ tool for specific purposes. Method submitters can be developers of the method, applicants or sponsors. They can range from academia, research institutions to industries. Following the assessment, EFSA should publish a report with indications of e.g., the capacities/ uses that can be verified based on the provided information (i.e., confirmed fit-for-purpose qualification), recommendations for developers and users, and indications for reporting experimental designs and results.

**The overarching aim of the proposed qualification system is to speed up the regulatory use of appropriate NAMs for risk assessments of NMs in the food and feed sector and to build confidence for their routine application.** The qualification

[10]www.gov4nano.eu
[11]www.diagonalproject.eu
[12]https://www.efsa.europa.eu/en/art36grants/article36/gpefsamese202201-nams4nano-integration-new-approach-methodologies-results

system itself can be regarded as a practical tool, which allows to a) evaluate NAMs regarding their suitability for a specific context-of-use, b) support the method development process (i.e. research/ optimisation/ validation) of NAMs and c) ultimately facilitate their regulatory implementation for risk assessments of NMs in the food and feed sector. As an outcome, a certain NAM can then be applied for the "qualified use" without the need to reconfirm/ reassess its suitability for each risk assessment. Thereby, this system can contribute to build confidence in NAM-based data for risk assessments. As all qualified methods will be published, the system also generates awareness and accelerates the widespread use of such qualified NAMs. Also, in the long term (at least for certain cases) the qualification system can support and enhance the process of method validation to establish new OECD TGs.

It should be kept in mind that, even though every method requires some kind of validation to demonstrate its scientific validity (i.e. reliability and relevance), not all methods may require a rigorous validation process as needed for establishment of OECD TGs. Furthermore, some methods are not of immediate interest for international harmonisation as they are not relevant for international regulatory use. The EFSA Guidance on Nano-RA (EFSA Scientific Committee, 2021a) specifically suggests a toxicity screening as a first step, involving assays for endpoints such as intestinal uptake/ crossing, cytotoxicity, reactivity/ oxidative stress or (pro-)inflammatory potential, which are non-classical endpoints that represent important KEs of NM toxicity. Such NAMs are not intended for stand-alone application but IATAs are foreseen to integrate the obtained NAM-data along with human and animal data. Nevertheless, the data obtained from such screening tests could be significant as it may or may not trigger higher tier *in vivo* toxicity tests. For such NAMs qualification can serve as a valuable endorsement tool. It is important that these NAMs are demonstrated to be scientifically valid as they are critical in tiered approach for NM risk assessment to minimise animal tests.

The objective of this report is to propose a generic framework for a fit-for-purpose qualification system for NAMs to be applied for NMs in the food and feed sector. Importantly, this report is only an interim version which should mainly serve as a basis to stipulate a broader discussion with EFSA, EU Member States, relevant organisations and stakeholders.

# 2 Data and Methodologies

## 2.1 Data Sources

The foundation of the present proposal are the existing qualification systems of US FDA and EMA. However, they operate in a different context, namely US FDA for Drug Development Tools and EMA for research and development into pharmaceuticals. They are not primarily intended for safety assessments. Therefore, they may only serve as examples for how to set-up the system and define its scope, aims and limitations. For instance, the FDA guidance for method qualification provides some important insights into the general requirements and the underlying process (FDA, 2023). However, additional considerations are needed for a NAM qualification system that would be appropriate for the food and feed sector.

The qualification system described in this document is aimed at NAMs that shall be applied in IATAs to support risk assessment of NMs in the food and feed sector. Therefore, we closely follow the EFSA Guidance on Nano-RA (EFSA Scientific Committee, 2021a). OECD GD 211 (OECD, 2014) was identified as the key publication for describing non-guideline cell-based methods. We also considered the more detailed ToxTemp template for describing non-guideline methods, which was established in the EU ToxRisk project to provide more clarity and guidance in support of the GD 211 (Krebs et al., 2019). In addition, OECD GIVIMP (2018a) and EURL-ECVAM test submission template[13] are considered as other overarching guidance documents providing useful information on the description of methods being relevant for regulatory acceptance. The OECD GLP principles (1998) were also taken into account as they include important considerations on the organisational process and the conditions under which non-clinical health and environmental safety studies should be planned, performed, monitored, recorded, archived and reported.

We identified three key approaches for assessing method readiness (Table 2), which were developed to support method validation for establishing OECD TGs. They include important criteria, most of which are equally relevant for evaluating method readiness in the context of method qualification.

Table 2: Overview of three identified key approaches to evaluate method readiness

| | Source | Main objective | Limitations |
|---|---|---|---|
| 1 | Bal-Price et al., 2018 | - evaluation of method readiness for DNT to facilitate validation/ regulatory implementation | - covers *in vitro* methods only<br>- no specific consideration for the food and feed sector<br>- no specific consideration for NMs |
| 2 | ReadED Test/ Pepper (Crouzet et al., 2023) | - evaluation of method readiness for ED to facilitate validation/ regulatory implementation | - covers *in vitro* methods and simple whole organism models (e.g., zebrafish) |

[13]https://joint-research-centre.ec.europa.eu/document/download/34e99b98-cd37-40d2-9779-382921ca04b6_en?filename=TPF%20Test%20Presubmission%20Form.docx

| | Source | Main objective | Limitations |
|---|---|---|---|
| | | | - no specific consideration for the food and feed sector<br>- no specific consideration for NMs |
| 3 | TRAAC (Shandilya et al., 2023) | - evaluation of method readiness for nano-specific methods | - very comprehensive/ complex criteria catalogue<br>- completeness assessment mainly from REACH perspective |

## 2.2 Methodologies

Firstly, we evaluated the existing qualification systems (US FDA and EMA), which served as examples to prototype our system although they do not qualify NAMs for safety assessment. In addition, we considered a) how risk assessment in the food and feed sector is done, in particular for NMs where routinely data from non-guideline studies is utilised, along with b) the known challenges and limitations for applying NAMs, especially non-guideline NAMs, in risk assessments.

We also evaluated the scope of such a system in order to make a suitable proposal how broadly the qualification system shall be initially established. Here, we took into account the unique opportunities for NAMs for nano risk assessments.

Next, we compared existing guidance for harmonising the description of non-guideline methods and existing key approaches to evaluate method readiness. In particular we compared the underlying criteria and how they are structured into overarching sections and sub-sections.

Our aim is to harmonise the requirements and guidance for submitting NAMs in applications such that the most relevant aspects for qualification will be covered. We identified that providing a detailed method description is the most crucial aspect when non-guideline methods are applied in risk assessment. Thus, we put emphasis on defining appropriate criteria for "detailed method description". In addition, we considered necessary elements for demonstrating the reliability and the relevance of a method but keeping in mind that there is little added value when duplicating all the criteria that are deemed necessary for method validation in the context of establishing OECD TGs. We therefore aimed to identify the absolute essential criteria to propose a much less rigorous process compared to OECD TGs.

Finally, we applied our approach to one selected method for two separate scenarios.

# 3 Assessment/Results

## 3.1 NAMs Qualification System for nanomaterials (NMs) in the Food and Feed sector

### Objectives of the NAMs Qualification System

Here we propose a **generic framework** for a fit-for-purpose NAMs qualification system to be applied for NMs in the food and feed sector with a clearly structured process considering the phases:

i) Submission Phase (proposal describing the NAMs by method submitters);
ii) Evaluation Phase (proposal being evaluated by EFSA experts);
iii) Outcome Phase (possible outcomes with consequences and next steps for method submitters, evaluating experts and risk assessors).

In addition, we aim to provide clear guidance for these three phases and their individual steps. Firstly, the guidance shall assist method submitters:

i) how to structure their submission for each stage (details provided later);
ii) which type of information and evidence is required at which stage;
iii) what will be the criteria for the evaluation of the NAMs;
iv) what is the expected duration of the process;
v) what are the possible outcomes;
vi) what are the consequences of a particular outcome/ decision.

Secondly, the guidance shall assist the experts on how to evaluate the submissions and prepare the outcome/decision reports. Thirdly, the guidance facilitates the dialogue between method submitters and evaluating experts.

The clear description of the process and the requirements should lower the barriers for submitting proposals. Also, the qualification system and in particular the guidance will serve as a tool to streamline and support the development of NAMs to speed-up their regulatory implementation for a well-defined context-of-use. Overall, the information requirement for submitting a NAM for qualification will be lower compared to those required in validation to establish OECD TGs. This may encourage a high(er) number of submissions. Importantly, the process is structured in a manner to identify and support promising NAMs already early in method development (i.e. NAMs that may not yet fulfill all the requirements). In such case, specific advice can be provided to support their further method development such that the NAM can reach qualification with the most efficient use of resources.

### Scope of the NAMs Qualification System for NMs

The EFSA Guidance on Nano-RA (EFSA Scientific Committee, 2021a) expects to consider data that are obtained using non-guideline methods. Therefore, the area of NM risk assessment would be in particular well suited to initially establish a NAM qualification system to speed-up the regulatory use of non-guideline methods in risk assessment provided they have a well-defined context-of-use and are demonstrated to be scientifically valid (i.e. reliable and

relevant) for this use. The following reasons provide additional explanations to clarify the need for qualification system specifically for NMs.

a) The existing OECD TGs are in general not suitable as such for NMs but they require adaptations, which is an ongoing process.
b) The particulate nature of the test items demands specific considerations. Firstly, NMs require a highly sophisticated physicochemical characterisation not only of the pristine material. Secondly, complex transformations that NMs will undergo in a biological environment such as (partial) agglomeration and dissolution may need to be considered as they may impact the assessments. Finally, issues such as dispersion stability, dosimetry or assay interferences have to be addressed to obtain meaningful test results. These issues render amendment of existing and establishment of new OECD TGs for NMs more challenging and time consuming than those for conventional chemicals.
c) The risk assessment of NMs is different compared to chemicals. Usually there are already existing data for their non-nano counterparts but it is unclear to what extent the existing data would also adequately cover all or some of their nano-variants. Another regulatory relevant question is which nano-variants are similar enough to be assessed jointly (e.g., by grouping and/or read-across).
d) Some essential elements of the NM risk assessment (see Figure 1) are technically easier to implement *in vitro.* Examples are the intestinal barrier crossing of particles or their cellular uptake.

Taken together these arguments demonstrate that, firstly NM risk assessment urgently requires NAMs. Secondly, the qualification system provides a unique opportunity to explore new ways to speed-up their regulatory implementation in specific sectors (here for the food and feed sector). Nanotechnology as an area with high innovation already has well established dialogues between the different stakeholders (public research institutes, industry, civil society and regulators) thereby ensuring that new concepts receive attention early on in their development stage. **Therefore, we propose to initially set-up and explore a NAM qualification system firstly to support NM risk assessment in the food and feed sector**.

Once this first qualification programme, which specifically targets "NAMs for nano", is successfully set up, EFSA may consider to establish other programmes to expand the scope into other relevant areas (e.g., risk assessment of chemicals, other endpoints or other contexts-of-use).

## Foreseen Context-of-Use of the qualified NAMs

Here, we propose a first EFSA qualification programme "NAMs for nano" to target the initial steps described in the EFSA Guidance on Nano-RA (EFSA Scientific Committee, 2021a), as schematically summarised in Figure 1. Hence, this programme should cover:

i) **NAMs for NM physicochemical characterisation**;
ii) **NAMs for characterisation of NM in relevant biological fluids** (e.g., to assess solubility, dissolution/degradation, other relevant particle transformations that have an impact on the assessments);

iii) **NAMs for toxicity screening** (e.g., covering in particular intestinal uptake/ crossing, genotoxicity, cytotoxicity, reactivity/ ox. stress, (pro-)inflammatory responses and barrier integrity, including multi-parametric methods and HTS, that could inform on key events in adverse outcome pathways).

## The Proposed Process

The proposed process is a sequential approach (Figure 9) that contains elements from existing qualification systems (US FDA and EMA). Our system is designed in three stages (Figure 9) to "guide" and "accompany" the test method development such that the scientific validity of promising NAMs can be demonstrated in a resource-efficient manner. Thereby, the method submitter has flexibility in submitting NAMs at any stage of method development. It is advisable to evaluate the "method readiness" at different stages of method development, in particular between the stages of the qualification system.

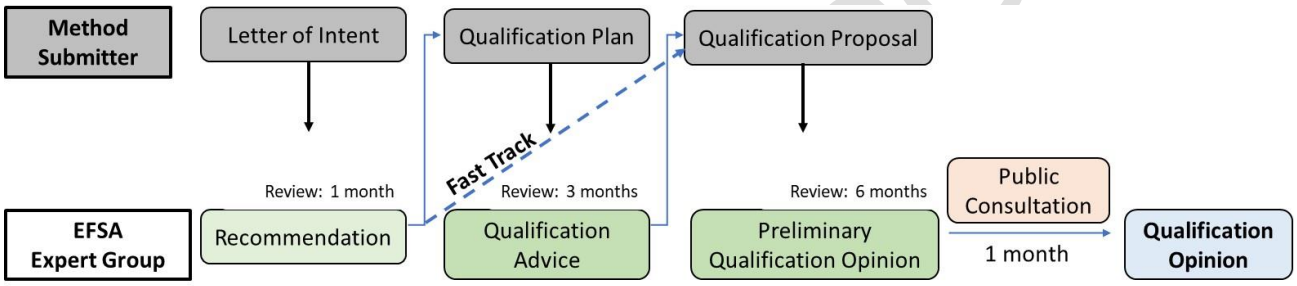


Figure 9: The suggested process for a NAMs qualification system for NMs in the food and feed sector

## Submission Phase

The submitters of test methods are encouraged to contact EFSA as early as possible to initiate a dialogue and to seek general advice, preferably even before the submission of a letter of intent (LoI). Below are the details on information requirements for the respective stages of the qualification system.

Stage 1: Letter of intent (LoI) - initial submission for first formal evaluation

The LoI is to seek a first formal evaluation of the NAM by the EFSA Expert Group and to obtain specific advice on how to proceed with method development. Hence, only minimum information on the method is requested from the submitter in this phase. The LoI should contain a) a summary of the method, some general information and the intended application, b) the scientific rationale, c) key details on the method description and d) an overview on which data sets exist to provide scientific evidence. An initial self-evaluation of "method readiness" is also expected as explained in the next chapter. The LoI is suggested to be evaluated within max. 1 month.

NAMs with a suitable readiness level (as explained in the next section) can be recommended for a "fast track" directly to a qualification proposal (skipping the

interim submission of a qualification plan). However, many NAMs might not yet be well enough developed to directly pass to qualification and are expected to follow the sequential approach outlined in Figure 9.

In any case, the method submitter will receive a letter from EFSA, which contains specific recommendation(s) on how to prepare and submit a qualification plan and/or advice(s) on how to develop the method.

Stage 2: Qualification Plan - interim submission for seeking qualification advice

The qualification plan is an interim submission for NAMs which are not yet "ready enough" for qualification. They can benefit from specific qualification advice from the EFSA Expert Group, including a letter of support (if needed). The method submitter is expected to provide sufficient details on method description along with all available data/ evidence to demonstrate its validity (i.e. reliability and relevance). It is also expected that an analysis of the remaining gaps for method development (in order to reach the qualification proposal stage) is done and that specific action items along with a time plan on how to address these gaps are proposed. The evaluation of qualification plan will take more time (suggested max. 3 months). A favourable outcome will be a qualification advice which will include specific recommendations for completing the next step. If useful, EFSA may also provide a letter of support, which could support the remaining phase of test method development, e.g. to apply for specific funding.

Stage 3: Qualification proposal - final submission aiming at a qualification opinion

The third and final stage of the proposed qualification system is the qualification proposal, which (if successful) is expected to result in a qualification opinion. It contains the full description of the method along with all detailed data that are necessary to demonstrate its validity (reliability and relevance) for its intended purpose. Thus, at this stage the submitter needs to demonstrate the appropriate method readiness (for regulatory use), underpinned with all the needed data/ evidence for evaluation (i.e., detailed data sets, including the raw data, analysed/interpreted data considering reference chemicals/ benchmarks for an appropriate number of test chemicals). Furthermore, at this stage the data obtained by the NAM should be already integrated with existing knowledge (e.g., *in vivo* data), which is an essential element to evaluate its relevance for the intended context-of-use. It is suggested to evaluate full qualification proposals within max. 6 months. A favourable outcome will be a preliminary qualification opinion that is published for public consultation. Later, the final qualification opinion will be published by EFSA.

Technically, we suggest that all submissions should be made via a dedicated electronic portal, which EFSA may consider to establish for this purpose. If technically feasible, the existing portal could be adapted. An electronic portal could ensure the technical completeness of all submissions according to the different stages of the process, the sequential order with

appropriate considerations of the recommendations (obtained at prior stages) and finally ensure a proper data/ meta-data format, which is useful for evaluation of the supporting scientific evidence. Until the portal is established, methods can be submitted as papers (following the guidance provided below). For paper submissions, it might be advisable to prepare templates for each stage or at least to provide examples. In order to raise awareness and to bundle all relevant information, EFSA may consider to set up a website for NAM qualification.

### Evaluation Phase

We suggest the following time frames as maximum for evaluations: 1 month for LoI, 3 months for qualification plans and 6 months for qualification proposals. The evaluations shall be conducted by a dedicated expert group, which can be established as an EFSA Working Group specifically dealing with NAMs qualification. Importantly, the expert group needs different expertise. We suggest, it shall comprise of:

- EFSA core staff (2).
- Validation experts (2).
- Technical experts for the specific area/ type of NAM (2)
    - EFSA should therefore nominate different experts for **physicochemical characterisation of pristine NMs** (2), **for NM characterisation in relevant biological fluids** (2), for ***in vitro* toxicity screening** (2-4) and for ***in silico* NAMs for nano** (2-3), even though the latter are currently only emerging.
    - depending on which NAM (type) is going to be evaluated not all experts need to participate but only those that have the required specific technical/ methodological expertise.
- EFSA experts for risk assessment depending on the intended application (2).

Depending on the method under evaluation, additional experts (2) may be invited, e.g., statistical experts or experts for high throughput approaches. Overall, the group could consist about 20 experts but only 8-10 will be needed during a specific evaluation.

During the whole process, EFSA can exchange with the method submitter to clarify additional questions.

### Possible outcomes and consequences

After each stage, the outcome can be either favourable or not. A favourable decision (accept) is needed to proceed to the next stage. This ensures that only methods that are developed well enough enter the next stage of qualification. If not favourable, further method development might be needed to improve the method to pass this step. Recommendations will be included that help to improve the methods. After further development (optimisation) the methods can be re-submitted at the same stage.

A favourable decision at stage 3 will be a preliminary qualification opinion, which will be published for public consultation (open for comments for 1 month). At the end of the process, a final qualification opinion will be published by EFSA, to generate awareness on application and usability of the method. Please note that additional time is needed after the public

consultation has ended, depending on the comments received that may need to be addressed by the method submitter, upon consultation with EFSA experts. Thus, in total up to 3-6 months might be expected before a final qualification opinion can be published on the EFSA website.

Importantly, seeking qualification of a NAM is voluntary. NAMs which have not been qualified may still be applied in risk assessments. However, they still require time-intensive case-by-case evaluation of their scientific basis for each submission.

## Qualification and OECD TGs: Commonalities and Differences

Qualification of a NAM and establishment of an OECD TG are two distinct processes that share commonalities but have differences, too (Figure 10).

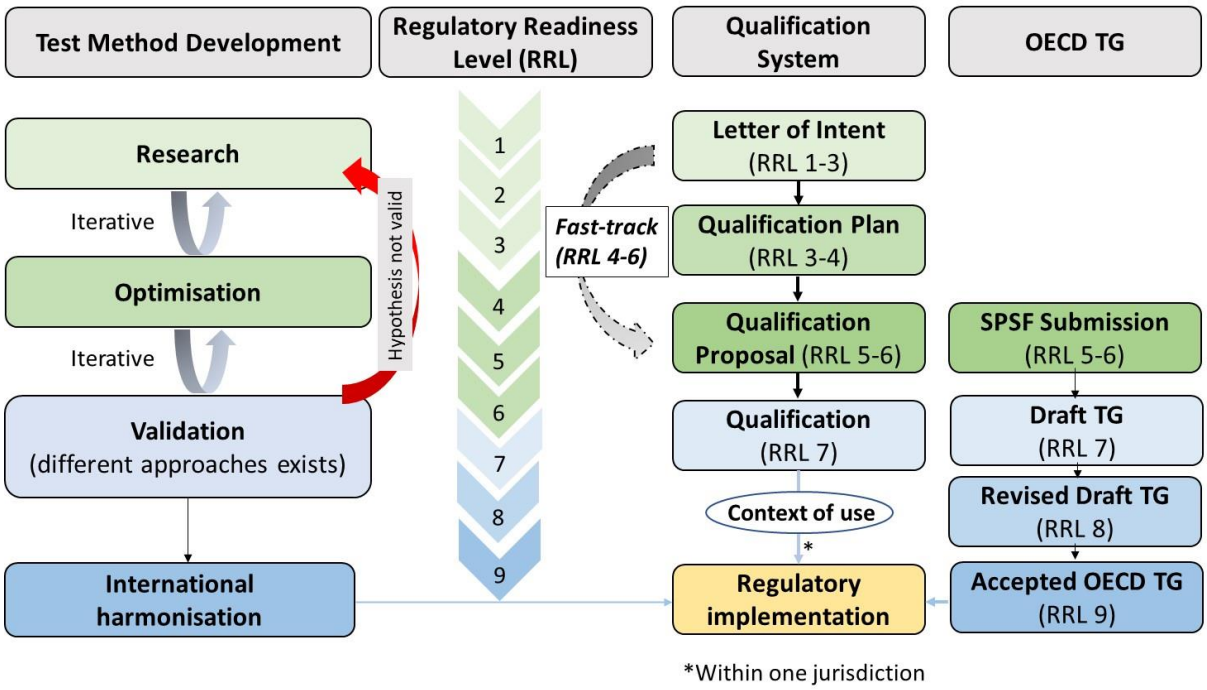


Figure 10: Overview on qualification of a NAM and establishment of an OECD TG and how both processes align with respect to test method development and regulatory implementation

As depicted in Figure 10, for both processes, a test method must have reached an adequate level of "readiness". In this context, we propose to introduce the concept of "regulatory readiness levels (RRLs)" for evaluation of "method readiness" to assess the test method precisely for it "regulatory readiness".

## Regulatory Readiness Levels (RRLs)

Overall, we suggest distinguishing between three categories of RRLs, namely i) research (RRL 1-3), ii) optimisation/ validation (RRL 4-6) and iii) regulatory implementation (RRL 7-9). During research stage (RRL 1-3) a test method is initially set up or defined. Already at this stage method submitters are encouraged to send a LoI to receive specific advice by EFSA

experts concerning the further method development for regulatory use. During optimisation/ validation (RRL 4-6), the method is further developed for an eventual application in the context of risk assessment. This includes the optimisation of the method description, demonstrating the scientific validity (validation) and preparing the method for regulatory acceptance. Regulatory implementation (RRL 7-9) typically begins with international harmonisation by submitting an SPSF with the aim to establish an OECD TG. As explained earlier, ideally, the methods are already validated at this stage (RRL 6). However, not all methods have already reached this stage when a SPSF is submitted. In such cases (RRL 5), the establishment of an OECD TG will take significantly more time and validation studies (including ILCs) are then conducted at the OECD level. Thus, in reality, often test methods are at RRL 5-6 when a SPSF is submitted. As already emphasised several times, the requirements for submitting a qualification proposal are lower (RRL 5) than those needed for establishing an OECD TG. It is assumed that most NAMs will not be validated (by ILCs) when they are submitted for qualification. Table 3 summarises the different RRL sorted into the three categories along with their overall draft criteria.

Table 3: Proposal for defining global "regulatory readiness level, RRL" and draft overall criteria for each RRL

| RRL | Category | Overall (draft) criteria |
|---|---|---|
| 1 | Research | ✓ novel test principle described, first scientific publication(s) |
| 2 | | ✓ method applied for several chemicals/ NMs,<br>✓ initial SOP(s), initial data on reliability (i.e., within-lab) |
| 3 | | ✓ first comprehensive data sets (possibly also from other labs)<br>✓ **detailed SOP(s)**<br>✓ ***method readiness evaluation: method scientifically plausible*** |
| 4 | Optimisation | ✓ remaining gaps evaluted, plan to close gaps established<br>✓ consider submitting a **Qualification Plan** |
| 5 | | ✓ (initial) **regulatory use defined,** first regulatory applications (e.g. in a risk assessment case study), transferability of NAM<br>✓ ***method readiness evaluation: method description completed (detailed harmonized SOPs available)***<br>✓ prepare validation study<br>✓ consider submitting a **Qualification Proposal** |
| 6 | Validation | ✓ conduct validation study<br>✓ consider submitting **SPSF to OECD** |
| 7 | Regulatory Implementation | ✓ **NAM "qualified"** (for one sector only) and/or<br>✓ NAM described in a **draft OECD TG** |
| 8 | | ✓ Draft TG revised |
| 9 | | ✓ **OECD TG endorsed/accepted** |

One may consider to add a few more RRLs (as needed) as the regulatory implementation typically is not completed with an endorsed OECD TG (e.g. OHTs still have to be established afterwards). However, this is outside the scope of this document.

## Differences between OECD TGs and Qualification: Context-of-Use

The main difference of both processes, qualification of a NAM and establishing an OECD TG, is that both define two distinct routes for regulatory implementation of test methods with different implications. **OECD TGs aim for consensus building and for international application of a test method, in line with the principles of MAD.** Each OECD TG addresses a specific and clearly defined regulatory need that often covers a wide(r) range of regulatory frameworks operating in different sectors (e.g., biocides, chemicals, food and feed) and this regulatory need is of interest in several countries. Establishing an OECD TG requires consensus among all OECD member states.
**In contrast, qualification of a NAM is a process that is based only on an expert evaluation and aims for implementation only in one legal framework**. Qualification allows for the use (or regulatory implementation) of a specific method only in one single sector (here: the food and feed sector), maybe even only for a specific sub-sector (e.g. only for pesticides or only for food contact materials) and specifically only for a well-defined (narrow) context-of-use (here: to specifically address the first steps for NM risk assessment as defined in the EFSA Guidance on Nano-RA (EFSA Scientific Committee, 2021a)). **Therefore, the primary distinction between them is their different context-of-use.**

Generally speaking, NAM results can be applied for different purposes in the context of risk assessments, covering e.g., a) prioritisation and screening, b) hazard/ exposure identification and/or characterisation, c) supporting grouping and read-across or e) providing arguments to waive specific tests. Furthermore, results can be used qualitatively or quantitatively. Existing OECD TGs serve specific regulatory data requirements. They provide the basis for addressing clearly defined regulatory relevant endpoints such as e.g. for hazard identification according to the UN Globally Harmonised System of classification. Some OECD TGs are based on stand-alone NAMs that can directly predict an adverse effect (e.g., those for skin or eye irritation). Other OECD TGs require the use of integrated approaches composed of several (non stand-alone) NAMs, which for instance address KEs of a known AOPs that are embedded in well-defined, refined (and OECD-endorsed) IATAs, often along with specific data interpretation procedures as exemplified by the DA to predict skin sensitisation. Thus, NAMs in OECD TGs are intended for predicting adverse outcomes to allow for regulatory decision-making.

The proposed NAM qualification system has a context-of-use that is distinct compared to establishing OECD TG. The biological NAMs, which can be submitted for qualification, are not stand-alone NAMs. They are foreseen to be applied in IATAs for initial screening and their data has to be integrated along with human and animal data, following a tiered process for risk assessment of NMs as requested by the EFSA Guidance on Nano-RA (EFSA Scientific Committee, 2021a). Importantly, most of the toxicological tests requested (i.e., cytotoxicity, reactivity/ ox. stress, (pro-)inflammatory responses and intestinal barrier impairment) are not even regarded conventional endpoints for hazard identification in risk assessment. Most of them represent KEs of NM toxicity. **Therefore, these methods are suitable only as screening methods**, rendering them mostly not of immediate interest in other EU or international regulatory frameworks (e.g., EU REACH) and consequently most of these NAMs are currently not prioritised for new OECD TGs. This renders them ideal for testing a qualification system as an alternative way to get such methods qualified for regulatory

implementation and to gain more experience in their application in a regulatory context. On the mid term, these experiences may be useful to facilitate the validation and further the international harmonisation at least for selected NAMs. Thus, over the longer term, qualification may also support the establishment of new OECD TGs.

### Possible long-term benefit: Qualification to support new OECD TGs

Qualification and validation/ international harmonisation to establish an OECD TG could be regarded as partially interlinked. In some cases, the qualification of a NAM could support its later validation/ international harmonisation by helping to "optimise" it for a particular regulatory application. A qualification system can thereby help (at least in some cases) to bridge the "implementation gap" between initial method establishment and its routine regulatory application. Specifically, qualification supports the following aspects:

a. clearly defining the regulatory need and (initial) intended use(s);
b. paving the way for risk assessments which integrate new NAM-generated data into existing data allowing to better understand the regulatory relevance of a specific method;
c. "prototyping" a NAM in an operational environment for well-defined regulatory questions;
d. yielding optimised SOPs (according to the information needs);
e. enabling trust building.

The latter aspect is due to increased awareness about qualified NAMs followed by increasing use of such qualified NAMs in risk assessments. The qualification process is geared towards supporting optimisation of promising NAMs to become "qualified" such that they could be transferred more easily to the late stage of the validation process.

### Practical application of the Qualification System

Overall, both processes describe different routes of regulatory implementation with different implications. The main difference lays in the "context-of-use concept". Figure 11 provides a schematic overview and gives some guidance for when qualification might be useful to consider.

Overall, two distinct "options" of application are foreseen for the qualification system:

i) Implement sector-specific considerations/ modifications in existing OECD TGs;
ii) Facilitate regulatory use of novel methodologies in a precisely defined context.

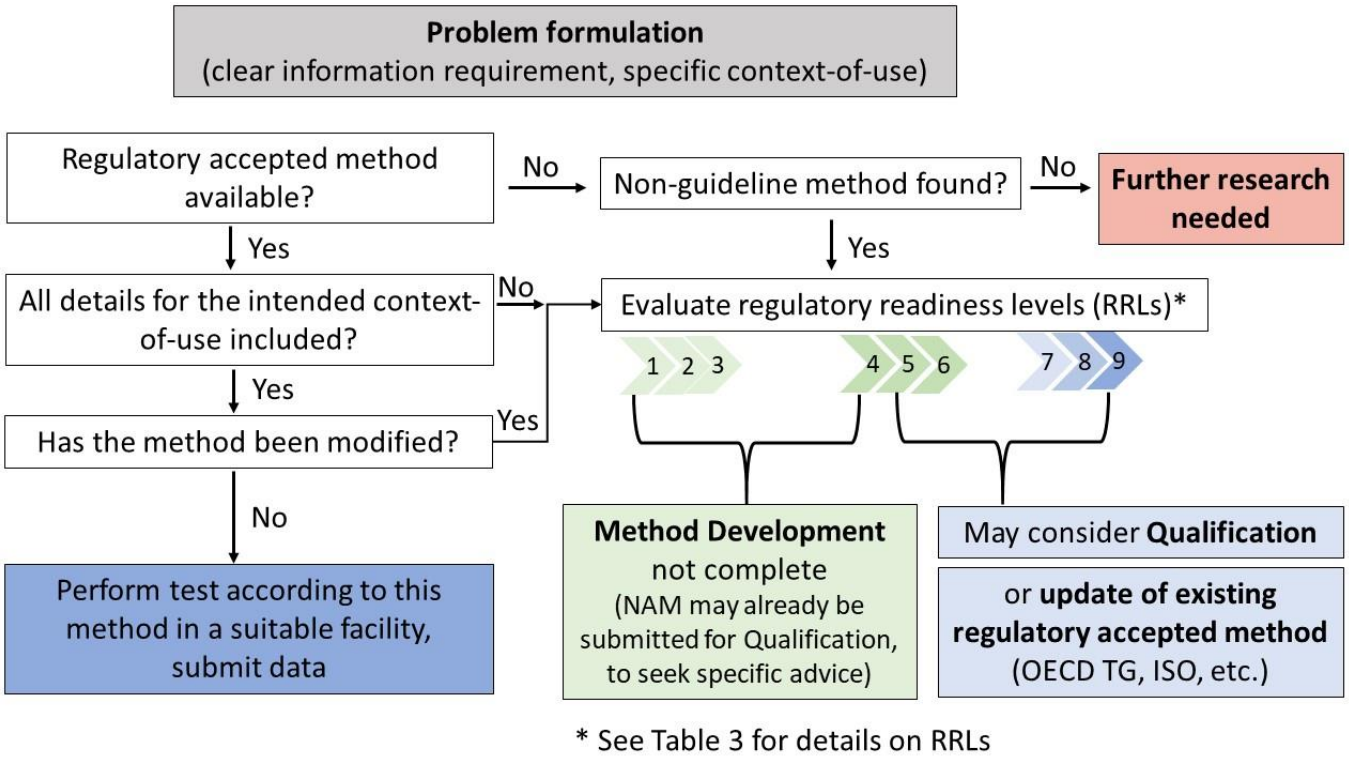


Figure 11: Overview and initial guidance for when to consider qualification

## Implement sector-specific considerations in existing OECD TGs

As explained earlier, OECD TGs are designed for international use within the MAD, usually they shall be applied in several sectors. Hence, they may not always consider all the particular details needed in one specific sector. In such cases, it might be feasible to include additional specifications that apply only for one particular sector via a qualification instead of aiming to formally adapt an existing OECD TG. An example for this type of application is particle size measurement by electron microscopy (EM). The OECD TG 125 (2023i) describes the measurement principles, applicability, prerequisites and influencing factor of EM applied for particle sizing in general terms, providing technical guidance largely supported by ILCs with relatively simple (reference) materials, such as near-spherical particles in an aqueous medium. This not necessarily implies that the RRL of the methodology described for EM in OECD TG 125 is sufficiently high for (more) complex cases involving various combinations of complex food matrices, particles compositions, size and shapes and particle numbers. Hence, in view of regulatory qualification of EM for analysis of (nano)particles in food products, additional method development and validation, and update of the guidance is necessary, depending on the complexity of the analysis (Figure 11). In this perspective additional specifications are, for example, in preparation by EFSA[14]. This document will provide practical instructions for reporting the results of EM analysis of materials possibly containing a fraction of small particles including nanoparticles in food and feed products. The objective is to support applicants during the reporting of an EM analysis, as well as risk assessors during the

[14] https://www.efsa.europa.eu/sites/default/files/2024-05/wg-nanotechnologies%20%281%29.pdf

evaluation of such data in view of applying EFSA guidance (EFSA Scientific Committee, 2021a; 2021b).

### Facilitate regulatory use of novel methodologies in a precisely defined context

The other type of application will be to facilitate the regulatory use of novel methodologies (i.e. those for which there are no OECD TGs) for specific and well-defined contexts-of-use. As explained in detail earlier, currently this will include only screening methods (mostly toxicity screening). In section 3.3. an example for this type of application is provided, namely the triculture model to evaluate uptake and intestinal barrier crossing of NPs and evaluate barrier integrity.

### Initial analysis of strengths, weaknesses, opportunities and threats (SWOT)

The following initial SWOT analysis (Figure 12) has been prepared as a basis for initiating discussion among experts and stakeholders on possible strengths, weaknesses, opportunities and threats of a NAM qualification system in the food and feed sector.

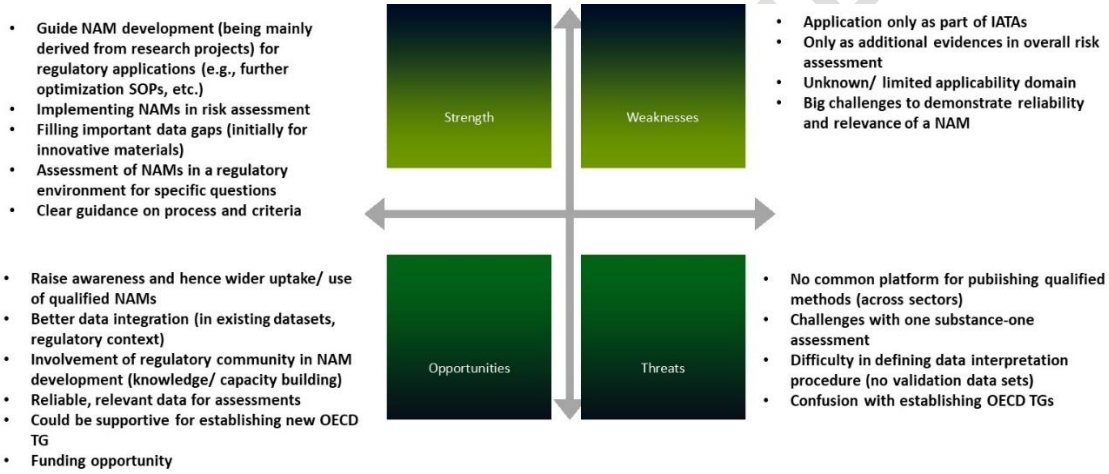


Figure 12: Initial SWOT analysis for a NAM qualification system in the food and feed sector

### Strengths

- The qualification system itself can be regarded a useful "tool" to guide the development of NAMs as test methods for regulatory purposes. It should be kept in mind that most of the NAMs for NMs are derived from research projects and are not yet validated. Therefore, they may need further work (e.g. further optimisation of SOPs) before their regulatory implementation.
- Implementing new technology (NAMs) in risk assessment. Already now non-guideline methods have to be used for NM risk assessment. However, once NAMs can become qualified the time-consuming and complex case-by-case evaluation of each method for each application will no longer be needed. This will speed-up the use of NAMs for risk assessment from case-by-case to routine level.

- Filling important data gaps for innovative material. The qualification of NAMs will increase confidence and awareness such that important data gaps, in particular for innovative materials (namely NM) can be filled in a timely manner.
- Clear guidance on process and criteria. Our qualification system provides clear guidance and informs on the process as well as on the criteria. This will support the submission of NAMs, their qualification and hence their later routine use.

### Weaknessess

- Application only as part of IATAs. Currently, stand-alone NAMs cannot be qualified. The only exception is to bring in sector-specific details/ modifications in an existing OECD TGs. Otherwise, the qualified NAMs are intended only to be used in tiered approaches (e.g. in IATAs).
- Added evidence only. As the NAMs are part of IATAs the obtained data will be considered as added evidence only (e.g. to make a decision whether the nano-specific considerations are already adequately covered in the existing assessments or which NMs can be assessed jointly).
- Unknown/ limited applicability domain. With the current state of the art for many NAMs for NMs it will be very challenging to determine or specify the applicability domains.
- Big challenges to demonstrate reliability and relevance of a NAM. As the vast majority of NAMs for NMs is not yet validated, it is currently very challenging to convincingly demonstrate their reliability and relevance, in particular for the biological screening NAMs. Currently, the "true" outcome for many NMs is unknown.

### Opportunities

- Wider uptake and use of qualified NAMs. The qualification system raises awaireness and thereby supports the wider use of NAMs for risk assessment.
- Better data integration (in existing datasets, regulatory context). As the NAMs are qualified it will be easier to use them for risk assessment. Their more widespread use will generate more data such that the data integration will be accelerated and will be potentially also improved by gaining more experience.
- Involvement of regulatory community (knowledge/ capacity building). Our system facilitates the dialogue between stakeholders and importantly involves the regulatory community in the NAM development. This will benefit both sides. Firstly, it will increase the knowledge on NAMs in the regulatory community (resulting in a much better view on their possible benefits or remaining limitation). Secondly, it will also help the method submitters to better understand regulatory requirements such they will be in a much better position to optimise their SOPs and demonstate the scientific validity of their NAMs for specific contexts-of-use.
- Reliable, relevant data for assessments but also for supporting validation. Qualification will support a routine use of NAMs in risk assessment thereby relevant data sets are growing (with each application), which on the medium or longer time scale may support validation and international harmonisation (for selected NAMs).
- Funding opportunity. The Qualification system can be a platform that also supports method developers to get funding. Promising NAMs are identified early and method submitters may receive a letter of support.

### Threats

- No common platform for qualified methods (across sectors). As currently several different qualification systems exist, which are not connected, there is no common platform where the qualified methods are published.
- Challenges with one substance-one assessment. Initially, the establishment of a NAMs qualification system only in the food and feed sector may be regarded as a threat for one substance-one assessment (however, on the longer perspective other sectors may benefit from the experience and even establishing new OECD TGs might be supported in several cases).
- Difficulty in defining data interpretation procedure (no validation data sets). It could be that initially qualification will be hampered as "validation data sets" are not available for most NAMs.
- Confusion with OECD TGs. Qualification applies key principles of validation – however, qualification and establishing OECD TGs are two distinct processes. Still, several stakeholders might be confused.

## 3.2 Guidance for NAM submission and evaluation

### Initial considerations

Non-guideline methods are already commonly used for NM risk assessment in the food and feed sector. Such cases need complex and time-consuming case-by-case evaluations. Currently, the non-guideline methods are mainly toxicity screening methods (as elaborated before). Therefore, in the following we will concentrate on biological NAMs, also because there are plenty of existing approaches (e.g. to evaluate the readiness of such methods). However, initial similar considerations for *in silico* and analytical (*in chemico*) NAMs are already included in the appendix. Importantly, the criteria and their guidance for biological NAMs are in alignment to OECD GD 211 (OECD, 2014) and ToxTemp (Krebs et al., 2019).

As elaborated before, we aim to circumvent case-by-case evaluations of the methods and "qualify" those that seem suitable (i.e. well enough developed, RRL 3-5) for their intended uses. In order to identify those that are well enough developed it is essential to evaluate their "readiness". Hence, we selected three key approaches for evaluation of method readiness (Bal-Price et al., 2018; Crouzet et al., 2023; Shandilya et al., 2023). Firstly, we compared the overall structure of all selected documents concerning their main sections (and sub-sections in two approaches) which guide the user on how to compile the needed information, for i) describing non-guideline cell-based methods or for ii) evaluating their method readiness (Table 4). As shown in table 4, they differ significantly with respect to their overall structure.

Table 4: Comparing the overall sections of selected approaches (OECD GD 211, Krebs et al., 2019, Bal-Price et al., 2018, Crouzet et al., 2023, Shandilya et al., 2023) that guide the user on how to structure the needed information for describing non-guideline methods and for evaluating the method readiness. In this table a color-code is applied to align similar sections (as far as possible), and then reproduced in the tables below (5-9)

| **Method Description** | | **Evaluation of Method Readiness** | | |
|---|---|---|---|---|
| *OECD GD 211* | *ToxTemp (Krebs et al., 2019)* | *Bal-Price et al., 2018* | *Pepper (Crouzet et al., 2023* | *TRAAC (Shandilya et al., 2023)* |
| General Information | General Information | | | Transparency, Accessibility |
| Test Method Definition | Description of general features | Test System | Test Method Description | Completeness |
| | Definition of test system as used | - | - | |
| | Exposure Scheme & Endpoints | Exposure Scheme | - | |
| | Handling Details | Documentation/ SOP | SOPs | |
| | - | Main Endpoint(s) | - | |
| | - | Cytotoxicity | - | |
| | - | Test Method Controls | - | |
| | - | Test Benchmarks | - | |
| Data Interpretation | Data Management | Data Evaluation | Operational Readiness and Data Management | |
| | Prediction Model | Prediction Model | | |
| Test Method Performance | Validation Status | Robustness | Reproducibility and Transferability | Reliability |
| | Transferability | - | Historical Data | - |
| Potential Regulatory Applications | - | Testing Strategy | Relevance of the method | Applicability |
| - | - | Applicability Domain(s) | - | - |

We applied a color-code in Table 4 to align similar sections (as far as possible). Occasionally, the different approaches use different words for similar concepts, e.g., test method definition/ test method description, data interpretation/ data evaluation or test method performance/ reproducibility and transferability. Some approaches are minimal in a sense that they comprise of essential sections such as OECD GD211, which has only 5 sections, while other approaches are very elaborated such as Bal-Price et al. (2018), which defines 13 sections including ranking parameters, performance criteria and a scoring system. In approaches from Pepper and TRAAC, though the number of main sections is comprehensive, they include many sub-sections which also might be complicated to apply for interested stakeholders.

As visible from the color-code, they overall share similar conceptual thinking. We aimed to propose a system which is minimal (with respect to the main sections) but comprehensive.

Importantly, the attempt was to structure the sections so that the meaning and purpose of their concepts are immediately evident to the method submitter.

We propose the following overarching sections:

- General Information;
- NAM Description (Test Method Description), preferably as SOPs, including:
  a. Description of Set-up: all needed elements/ components (Part I)
  b. Description of Application: how to apply/ use the NAM (Part II)
  c. Description of Evaluation: outcome and how to analyse/ interpret it (Part III)
- Reliability and Relevance (both in the context of the regulatory application).

Each of these sections is further specified into sub-sections. Each sub-section, finally, contains several individual criteria. The information requirements then increase (LoI, Qualification Plan, Qualification Proposal). A schematic summary (without all the details) is provided in Table 5.

Table 5: Overview on which aspects have to be considered when submitting NAMs for qualification considering the different phases, Letter of Intent, Qualification Plan, Qualification Proposal (*- information can be submitted along with LoI for fast-track to qualification proposal)

| | | **Letter of Intent (LoI)** | **Qualification Plan** | **Qualification Proposal** |
|---|---|---|---|---|
| General Information | | Name, Summary… | Updates (version, if applicable) | Updates (version, if applicable) |
| Method Description | | General Overview of NAM | SOPs covering all phases (set-up, application, evaluation) | Update SOPs<br>• acceptance criteria,<br>• limitations<br>• guidance for troubleshooting<br><br>(this relates to all phases, set-up, application, evaluation)<br><br>Data Interpretation |
| | Biological Model<br><br>(Test System) | Cell Type, Source | Routine culture & application in assay | |
| | Assay/ Readout | Components, Endpoint(s) | Working range, calibration | |
| | Application | Not expected in LoI* | Preparation of test items, Exposure regime, Test concentrations, Sampling/ Readout times… | |
| | Controls | | Positive/ Negative controls | |
| | Evaluation | | Raw Data, Preprocessing, Data Analysis | |
| | Reference Chemicals/ Benchmarks | | Not expected in LoI* | True positives/ negatives? Benchmarks? Performance standards? |
| Relevance | | Context-of-use | Underlying scientific principle & plausibility | Comparison to gold standard method (if applicable)/ Data Integration |

| | Letter of Intent (LoI) | Qualification Plan | Qualification Proposal |
|---|---|---|---|
| | | | Orthogonal/ similar methods |
| Reliability | Not expected in LoI* | Within-lab reproducibility (over time, different user) | Transferability |

The general information should contain key information on a method (e.g., name, method summary) and disclose any proprietary item.

The most important section in our opinion is a proper and detailed "Test Method Description", which is in line with how currently non-guideline methods are applied in NM risk assessments in the food and feed sector. Currently, a time-consuming case-by-case evaluation of the method itself is required by the assessor and the assessment critically depends on the details of the method description to decide whether the data from it can be applied for a specific assessment. Hence, the detailed test method description (preferably provided as SOPs) is of utmost importance also for qualification. Importantly, the detailed method description needs to consider all the phases of how the method will be applied. It starts with i) the set-up of the method (including a description of all the elements of a method, here the applied biological model and its routine culture and the intended assay/ readout), ii) the description of all steps of the application phase (including e.g. how to prepare the test items, how to cultivate the model specifically for the test, how to establish the tested concentration range, how to check for interferences of NMs) and finally iii) the evaluation phase (e.g. what are the raw data, how to deal with them/ any needed preprocessing, how to analyse them, how to interpret the obtained result).

In addition, it is equally important to demonstrate the scientific validity of a method, even if this might be challenging for most of the NAMs that are still under development for NMs. In many cases, for instance, larger ILCs are lacking. Anyhow, as explained before, scientific validity must include an assessment of the reliability and the relevance (both in the context-of-use) of the NAM.

In the following chapters, we introduce the sections in detail along with the sub-sections and their readiness criteria. Importantly, the level of expected details increases from an initial submission of a Letter of Intent (stage 1) to the final qualification proposal (stage 3).

## General Information

This section has no sub-sections and is not intended for evaluation. The general information provided here is to introduce the NAM to the EFSA Expert Evaluation Group (Table 6). It is required in all cases (i.e., for all stages: LoI, qualification plan, qualification proposal) and for all methods (i.e., *in silico*, analytical, biological).

It should be noted that this section should be re-visited for all submissions as some details might have changed. For instance, over the course of the qualification process it can be foreseen that the NAM may be updated such that several SOP versions exist (which may vastly differ from those of the initial submission). It is important to trace the developmental process of a method to understand if SOPs exist in several versions or with different names.

It must be clearly indicated in each version what was updated as parts of the associated evidence (data) might have been obtained with a previous (i.e. slightly different) SOP version.

In addition, the method summary may need to be revised as more information and knowledge becomes available. Finally, the developmental status is expected to be updated with each subsequent submission.

Table 6: General Information that always has to be submitted for all types of NAMs

| **Criterion** | **Guidance** |
|---|---|
| Name of NAM | Short, descriptive title |
| Synonyms | Other names used for a method |
| Version number | Document the process (if applicable)<br>Explain what has been updated in which version. |
| Summary | Max. 250 words, including all key features of the method (e.g., test system, measured endpoints, relevance, proposed context-of-use) |
| Intended Use | Please describe as detailed as possible the intended use |
| Date | XXX |
| Author(s) | XXX |
| Affiliation(s) | XXX |
| Contact Detail(s) | XXX |
| Type of NAM | *In silico* (PBK, QSAR);<br>*In chemico*; Analytical<br>Biological (low/ moderate/ high throughput) |
| Development Status | Self-evaluation of regulatory readiness level (RRL) |
| Abbreviations/ Definitions | If needed |
| Transparency on proprietary elements | Are all needed elements of the NAM available to all? This may influence the later application (e.g. if a cell model is available only in one lab, or if an analysis software is not yet disclosed). It may also influence evaluation (e.g. if not all datasets can be disclosed).<br>Elements of the test system may be protected by patents or any other means. This should be clearly stated here. |
| SOPs available/ published | Overview of all available SOPs (e.g. for routine cell culture, for test item preparation, for the assay, for data evaluation etc.)<br>Importantly, for a qualification proposal it is expected that all SOPs are published (e.g. scientific journal, website). |

A criterion on transparency was added to disclose any proprietary element, if existing. For instance, the ownership of a method needs to be provided and the accessibility of all elements (e.g., a cell model that is available only in one particular laboratory or an analysis software that is protected/ not commercialised). Proprietary elements may impact the future application of a method by others (maybe not all other laboratories will be able to apply the method) but may also impact its evaluation (e.g., if not all datasets can be disclosed).

In the following step-by-step guidance is given on each stage of submission (i.e. LoI, qualification plan and qualification proposal) to explain which information are needed at which stage. Please note that an excel table is provided in Annex I with all requirements in one document for a better overview. **Please keep in mind, that in general all available evidence on a method should be submitted.** Several optional criteria are included there.

They may not equally apply for all methods. Also, they are not considered essential but they will be helpful in addition (if existing). It is advised to use the information provided in the forthcoming chapters for orientation into the qualification system and then consult the excel table in Annex I for details.

### Information needed for Letter of Intent

In addition to the general information described above (Section 3.2.2) the following information should be included in the LoI (Table 7).

Table 7: Information being required for the LoI for cell-based biological NAMs

| Section | Sub-Section | Criterion | Guidance |
|---|---|---|---|
| **Test Method Description Set-Up (Part I)** | Biological Model | What is represented by the system? | ✓ Which organ (e.g. liver, kidney, lung) is represented?<br>✓ Which tissue (e.g. epithelial)?<br>✓ Important physiological/biochemical functions of the model? |
| | | Overview on all components | ✓ What is the model composed of (e.g. cell line, primary cells, iPS)?<br>✓ Species (human?)<br>✓ Other details (e.g. on donor: sex, age) |
| | | Source | ✓ Commercial supplier?<br>✓ How are the cells obtained? |
| | | Specific characteristics | ✓ Characterise source cells, mention differentiation if needed (e.g. iPSC)<br>✓ Cell identity (e.g. ordering number, passage number, purity of cells)<br>✓ Cell features (e.g. express/ lack specific markers)<br>✓ Co-cultures (percentage of each population)<br>✓ Organoids and microphysiological systems (MPS): ratio of cell types<br>✓ Is a specific cellular environment needed? |
| | Assay/ Readout | What is actually measured? | ✓ Which specific endpoints are measured?<br>✓ Indicate whether cytotoxicity is the primary endpoint (often cytotoxicity is the starting point for other cell-based tests), which other endpoints are included (e.g. genotoxicity, oxidative stress, inflammatory potential etc.)?<br>✓ Also describe here potential reference or normalization endpoints (e.g. often results are expressed in relation to another endpoint such cytotoxicity, protein content, housekeeping gene expression). |
| | | How is it measured? | ✓ Which measurement principle (e.g. mass spectrometry, absorbance, |

| Section | Sub-Section | Criterion | Guidance |
|---|---|---|---|
| | | | immunoblot, polymerase chain reaction…)? |
| | | Equipment needed? | ✓ Any custom-made instrument(s) or any specialised equipment(s) that is not readily available in standard labs |
| **Relevance (for context-of-use)** | | Biological Relevance? | ✓ Which biological process(es)/ Which toxicological events (e.g. oxidative stress, cell death) are reflected by the test method? |
| | | How does it fit in the regulatory framework? | ✓ Is it a screening assay?<br>✓ Which step in the EFSA Guidance on Nano-RA is it addressing? |

A LoI should always be submitted, for NAMs of all RRLs (1-6). It is the entry point before the submission of a qualification plan (RRL 3-4) or a qualification proposal (RRL 5-6). The aim is to initiate a formal dialogue with EFSA on the potential of the method and what kind of support it would require. The method submitter could get specific advice on the further method development. Methods that are evaluated to be already at RRL 5-6, can be recommended for the fast-track option (skipping the interim qualification plan stage) to final qualification proposal stage as outlined in Figure 9 (see section 3.1.4).

### Information needed for a Qualification Plan

In addition, to the general information (3.2.2) and the information provided already for the LoI (3.2.3), the following information is requested/expected for a qualification plan (Table 8). Please check if some of the information on the method, which was submitted in LoI earlier would require an update in the qualification plan.

Table 8: Information needed for a qualification plan for cell-based biological NAMs

| Section | Sub-Section | Criterion | Guidance |
|---|---|---|---|
| **Test Method Description (Overarching)** | SOPs | Detailed SOP(s) for all phases/ steps? | ✓ All elements of the method have to be described. Several SOPs may be required.<br>✓ Include all steps to describe the set-up of the NAM (e.g. routine cell culture).<br>✓ Describe all steps of the application phase (including test item preparation/ characterisation etc).<br>✓ The evaluation phase may need additional SOPs (depending on the NAM), otherwise details may also be included other SOPs. Anyhow, detailed information is needed on how to deal with raw data (preprocessing?), how to analyse them (curve fitting?) and how to interprete the outcome (what is positive, what is negative)? |
| **Test Method Description Set-Up (Part I)** | Biological Model | Characterisation of the model | ✓ Is used model well enough characterised? How?<br>✓ (e.g. any known causes of variability during routine culture), any critical |

| Section | Sub-Section | Criterion | Guidance |
|---|---|---|---|
| | | | consumables that influence the performance, any possible batch effects (e.g. fetal calf serum) |
| | Assay/ Readout | Qualitative/ Quantitative | ✓ Will the assay provide qualitative or quantitative information? |
| | | Calibration required? | ✓ Does the assay require a calibration and if so, how calibration should be performed? |
| | | Internal standards needed? | ✓ Are internal standards required? If so, which.<br>✓ Explain how they are used. |
| | | Negative controls | ✓ What chemicals/manipulations are used as negative controls? What is the expected data for such controls (signal/ uncertainty)?<br>✓ Highlight if there are compounds to be used for testing day-to-day test performance. |
| | | Positive controls | ✓ What chemicals are used as positive controls?<br>✓ Expected data on such controls (signal/ uncertainty).<br>✓ Is there *in vivo* reference data on the positive controls? Are *in vivo* relevant threshold concentrations known?<br>✓ Highlight if there are compounds to be used for testing day-to-day test performance. |
| | | Other controls | ✓ If any other controls should be included, indicate these compounds and the respective rationale for their use and the concentration selection. |
| **Test Method Description Application (Part II)** | Biological Model (as used) | Differentiation status for starting the assay? | ✓ Principles of the selected differentiation protocol (if needed), include a scheme/ graphical overview, describe all phases/steps, media, substrates (medium change/re-plating, medium additives, etc.). |
| | | Characteristics needed for starting the assay? | ✓ What are measures to decide before toxicity testing can be initiated (e.g any specific morphology, gene expression etc to be checked)<br>✓ Values (e.g. degree of differentiation/cell density) that should be reached/not be reached<br>✓ If two or more endpoints are important to decide describe if they are measured both in the same well, under same conditions, in parallel, or independently of each other?<br>✓ For imaging endpoints: Explain (in general) how to quantify or estimate, how many cells are imaged (roughly). |
| | | Specific cell culture format for assay? | ✓ Provide an exposure scheme (graphically, show timelines, addition of medium supplements and other compounds as needed, sampling, etc.) within the context of the overall cell culture scheme (e.g. |

| Section | Sub-Section | Criterion | Guidance |
|---|---|---|---|
| | | | freshly re-plated cells or confluent cells at start, certain coatings, etc.).<br>✓ Include medium changes, cell re-plating, whether compounds are re-added in cases of medium change, critical medium supplements, etc.<br>✓ Use of ALI system? |
| | | Specific cellular environment? | ✓ Specify if a specific environment is important for the assay, examples are specific glucose concentrations in medium, pH values that infere with readout, presence/ absence of serum during NM application … |
| | | Specific considerations for NM? | ✓ Are there any specific considerations for NMs?<br>✓ For instance, in organoids needed inside out?<br>✓ Capacity of particle uptake known (often a limiting factor in some cell models? |
| | Exposure Regime | How to select concentrations that are tested/ measured? | ✓ Often based on prior cytotoxicity tests but other ways are also possible (please specify)<br>✓ Does the test system have a particular apoptosis sensitivity or resistance? Is cytotoxicity hard to capture for minor cellular subpopulations?<br>✓ In multicellular systems, which cell population is the most sensitive? Are specific response markers known/ needed for each cell population?<br>✓ Are there issues with distinguishing slowed proliferation from cell death?<br>✓ For repeated/prolonged dosing: Is early death and compensatory growth considered? |
| | | How many test concentrations recommended? | ✓ How to define concentration range (e.g. always 1:10 serial dilutions or variable dilution factors)?<br>✓ How many concentrations should be tested? |
| | | How to select treatment times or measurement times? | ✓ Describe how the treatment/ measurement times were established? Were initially different time points tested?<br>✓ Indicate possibility/extent of repeated measures (over time) from same dish. |
| | | How many treatment times/ measurement times recommended? | ✓ Some assays may require several treatments.<br>✓ Some assays may allow for multiple measurements over time. |
| | | Specific considerations for NM? | ✓ How to test for NM interferences?<br>✓ Mention important details (e.g. if it is important to wash tissue models before assessing cell viability to remove NMs). |

| Section | Sub-Section | Criterion | Guidance |
|---|---|---|---|
| **Test Method Description Evaluation (Part III)** | Preparation of Test Item | Appropriate NM/ substance characterisation? | ✓ Nano-specific: Please specify how the NMs were characterised before the test? Which phys-chem. parameters are important?<br>✓ General: Please specify how substances should be characterised (e.g. level of purity, stability…) |
| | | How to prepare test item? | ✓ Nano-specific: how to prepare NM dispersions? How to characterise (e.g. for degree of agglomeration/ stability over time)?<br>✓ How are dilutions prepared? Which solvent? Is filtering used to obtain sterility? |
| | | In which medium/ dispersant is the test item prepared? | ✓ E.g. in cell culture medium (with/without fetal calf serum or serum replacement) or in another (biological) fluid?<br>✓ Which system for ALI exposure? |
| | | Equipment needed? | ✓ For example: sonicator |
| | Sampling | Specific sampling procedures? | ✓ Range of volume, instruments for dispensing, temperature, buffer, preparation of dilution, time-plan for pipetting, how to calculate compound concentrations, plate maps, addition of test compounds |
| | | Specific readout times? | ✓ If any (e.g. 30 min, 60 min) |
| | | Particle interferences appropriately considered? | ✓ What types of compounds are problematic, e.g. interference with analytical endpoint, low solubility, precipitation of medium components, etc.? |
| | Experimental Raw Data | Format of raw data? Unit(s)? | ✓ What is the data format?<br>✓ Raw data: give general explanation. Upload an exemplary file of raw data (e.g. Excel file as exported out of plate reader). |
| | | How to capture raw data? | ✓ If the file format is not proprietary or binary, include a template. This will help other users to provide their data in a similar way to the general data infrastructure. |
| | Metadata | Which metadata is needed? | ✓ What metadata are important?<br>✓ How are they linked to raw data? |
| | | Format of metadata? | ✓ Give example of the metadata file. |
| | | How to capture metadata? | ✓ How are metadata documented and stored (lab book, Excel files, left in machine, etc.)? |
| | Data Analysis | Any preprocessing? | ✓ Describe all preprocessing steps (e.g. background substraction, any normalisation etc) |
| | | How is data analysed? | ✓ How are data normally handled to obtain the overall test result (e.g. concentration response fitting using model X)? |
| | | Software? Workflow? | ✓ Specific software/workflow used for data evaluation (eg. Origin) |

| Section | Sub-Section | Criterion | Guidance |
|---|---|---|---|
| **Reliabiliy (for context-of-use)** | | Within-Lab Reproducibility (over time) | ✓ Monitor the performance over time (ideally collect a "historic dataset"). |
| | | Within-Lab Reproducibility (different user) | ✓ Evaluate performance if other users (of the same lab) apply the method. |
| **Relevance (for context-of-use)** | | Underlying scientific principle? | ✓ Indicate whether the test method has been or could be linked to an AOP (or several AOPs) and in which form (e.g. test of KE activation).<br>✓ What is the scientific rationale of this linkage? What is the relevance for an *in vivo* adverse outcome? |
| | | How firm is the evidence? | ✓ How firm is this evidence (e.g. is the AOP already endorsed by OECD or only drafted) |

## Information needed for a Qualification Proposal

In addition, to the general information (3.2.2) and the information provided before during the LoI stage (3.2.3) and the qualification plan stage (3.2.4), the following information is needed for a qualification proposal (Table 9). Please note that some of the information provided earlier may require an update. The qualification proposal covers aspects to establish the limitations of the method. Therefore, the method submitter needs to ask the following main questions: 1) what is the precision of the method, 2) under what conditions does the method work and not work, 3) within what range of accuracy does the method agree with the secondary standard and therefore measure reality?

Table 9: Information needed for a qualification proposal for cell-based biological NAMs

| Section | Sub-Section | Criterion | Guidance |
|---|---|---|---|
| **Test Method Description Overarching** | Acceptance Criteria | Biological Model (Test System) | ✓ Consider acceptance criteria for routine culture/ before test (e.g. pathogen-free, reach a specific confluence, show certain characteristics etc)<br>✓ Consider acceptance criteria during assay (e.g. specific functional parameters)<br>✓ Consider acceptance criteria that should be reached at the end of the assay<br>✓ Exclusion criteria are also important (features that should be absent) |
| | | Readout/ Assay | ✓ Consider acceptance criteria for assay (in which cases data shall not be used for analysis?) |
| | | Data Evaluation | ✓ Consider acceptance criteria for data evaluation |
| | Known Limitations | Biological Model (Test System) | ✓ Describe the limitations of the cell model (e.g. with respect to particle uptake, cellular redox-status if critical to readout etc) |

| Section | Sub-Section | Criterion | Guidance |
|---|---|---|---|
| | | Readout/ Assay | ✓ Any substances/ NMs that cannot be measured?<br>✓ Any difficult to measure substances/ NMs?<br>(e.g. low-density particles) |
| | | Test Item Preparation | ✓ For example, NMs that are difficult to disperse |
| | | Data Evaluation | ✓ Known limitations of data interpretation procedure? Both types of changes considered (up/down)? Or is model one-sided (e.g. toxicants leading only to a decrease/ increase)?<br>✓ Any thresholds, performance characteristics? |
| | Troubleshooting | Biological Model | ✓ Elaborate on possible issues, which could occur and how to address them |
| | | Readout/ Assay | |
| | | Test Item Preparation | |
| | | Data Evaluation | |
| **Test Method Description Set-Up (Part I)** | Biological Model Stability | Biological Model (Test System) | ✓ Elaborate on stability/ variability of the model (e.g. over several passages? Which are valid?)<br>✓ Any critical steps/ issues?<br>✓ For iPSC: How is controlled for pluripotency?<br>✓ For primary cells: Show stability and identity of supply; demonstrate stability of function (e.g. xenobiotic metabolism). |
| **Test Method Description Evaluation (Part III)** | Data Evaluation | Further preprocessing details: outliers? | ✓ How are outliers defined and handled?<br>✓ How are they documented?<br>✓ Provide the general frequency of outliers. |
| | | Data Interpretation | ✓ How to interpret results? When is the outcome considered positive/ negative?<br>✓ If a mathematical model (prediction model) is used: how was it set up (using which test set of chemicals to train the model; using probing with what kind of classifiers/statistical approaches)?<br>✓ Has the prediction model been tested (for how many substances/ NMs)? |
| | Reference chemicals | True Positives? | ✓ Known true positive substances/ NMs? |
| | | True Negatives? | ✓ Known true negative substances/ NMs? |
| | | Problematic substances/ NMs? | ✓ Are any problemetic substances/ NMs that would be known to give "borderline" results? |
| | | Proposals for reference/ benchmark NMs? | ✓ Suggestions for reference chemicals that later may play a role as performance standards?<br>✓ Any benchmark NM (in the absence of known true positives/ negatives they might help in meta-analysis, comparisons) |
| **Reliability (for context-of-use)** | | Transferability | ✓ Minimum one other lab should have applied the method. How comparable is outcome?<br>✓ Which amount of training is needed? |

| Section | Sub-Section | Criterion | Guidance |
|---|---|---|---|
| **Relevance (for context-of-use)** | | Compare to *in vivo*/ Integrate | ✓ Compare to *in vivo* data (if available)<br>✓ Integrate with all other available evidence |
| | | Human relevance? | ✓ Information on where the test system differs from the mimicked human tissue and which gaps of analogy need to be considered. |
| | | Data from orthogonal methods? | ✓ An orthogonal method is a method that addresses the same endpoint but uses a different measurement principle. |

## How to address Reliability

It might be foreseeable that for many NAMs for NMs not all the information on reliability are available already and often they are also not easy/ straightforward to obtain (e.g., transferability, between-laboratory variability). For many NAMs there are not enough resources to conduct larger ILCs as they are typically required to establish OECD TGs. Hence to confirm reliability of data from NAMs in the qualification system, three aspects are considered crucial. Firstly, the submitter needs to provide data on variability when using the NAM over time (in the same laboratory). It is considered important to establish a historical control data set. Secondly, the submitter needs to provide data on variability when different users in the same laboratory use the NAM. Thirdly, it is necessary to demonstrate transferability to minimum one other laboratory. The experiences of the lab transfer should be reported. It is important to understand whether specific training was needed, specific level of expertise was needed and also how comparable are the data.

## How to address Relevance for the Context-of-Use

It might be even more challenging to demonstrate the relevance for many NAMs for NMs for several reasons. Firstly, relevant *in vivo* data for *in vitro/ in vivo* comparisons are often lacking for many NMs. Secondly, there are no validated "reference methods" for comparison and typically there are no "known true positive" and "known true negative" NMs. Finally, comparison across different studies (i.e. meta-analysis) are also hampered by various aspects (many studies test only a small NM number, material characterisation is often incomplete and there are no commonly agreed benchmark materials etc.). Therefore, it is difficult to judge the relevance of a NAM and this also clearly hampers defining the applicability domain.

We suggest the focus on the following aspects to address relevance of the NAMs.

Firstly, the integration of the NAM-based data into existing other evidence (e.g. *in vivo* animal data or human data, if available). This is of central importance for evaluating the relevance of a particular NAM. However, as explained above, this will not be possible for all cases.

Secondly, the qualification system currently addresses only screening NAMs that are foreseen to be applied in IATAs. Hence, the submitter should elaborate on the scientific rationale behind the NAM. Which AOPs can be linked to this particular NAM? Which KE is actually tested? How firm is the available evidence for this AOP? Has it been endorsed already by the OECD? If not, is there any other evidence to support the AOP? For this, the submitter may consider testing

other KEs of the same AOP to provide more evidence. Also, the submitter could test other events (besides KEs of AOPs), which have a plausible linkage in the context of toxicity mechanism/ mode-of-action. For instance, if the NAM measures the release of a particular cytokine, the submitter may consider testing other endpoints in relation to a pro-inflammatory response (e.g. the upregulation of specific surface markers, the release of other cytokines, inflammasome activation etc). Furthermore, the submitter may also try to combine the nano-specific data (obtained with the particular NAM of interest) into a larger dataset (obtained with the same NAM but being applied to a larger set of conventional chemicals). The rationale behind the latter is that in general more data is available for conventional chemicals and some of this could provide arguments how the NM responses fit into the overall response patterns, provided that some toxicity mechanisms are shared. In such a case, one could aim to unravel similarities between specific NMs and specific chemicals and then try to find explanations on a similar mode-of-action (e.g., the release of toxic metals etc). Such elaborations could strengthen and support the scientific rationale.

Finally, relevance could also be supported by using a combination of orthogonal methods (Simon et al., 2023). Orthogonal methods are defined as "those that use different physical principles to measure the same property of the same sample with the goal of minimising method-specific biases and interferences" (Simon et al., 2023). Two measurements need to differ in at least one physical principle to be considered orthogonal. The actual goal is to determine the true value of a measurand taking into consideration biases and interferences inherent in the method (Simon et al., 2023). Typically, orthogonal methods are applied in analytical sciences in the contexts of reliability (determining the true value). Orthogonal measurements have been recommended already for particle size distributions, where different measurement techniques can be used such as dynamic light scattering (DLS) or EM. Each technique has a different physical principle to measure size and the measured value can vary (depending on the technique applied) which can complicate risk assessment (Simon et al., 2023). By comparing measurands using different methods, the data will be more informative and less biased (Simon et al., 2023). Methods cannot be orthogonal if they measure different properties such as size and composition (Simon et al., 2023). Furthermore, the methods are not orthogonal if they measure different size ranges (Simon et al., 2023).

The concept of orthogonality was also applied by van der Zalm et al. (2023) to 29 agrochemicals which were evaluated using *in vitro* and *ex vivo* methods for eye corrosion/irritation. The two methods evaluated comprised of Defined Approaches (DAs): EpiOcular EIT (OECD TG 492; EO) and Bovine Corneal Opacity and Permeability test (OECD TG 437; BCOP). The results from these two methods were compared to historical *in vivo* rabbit eye test in an orthogonal procedure. Lastly, it could be concluded that the two DAs (EO and BCOP) are fit –for-purpose and of human relevance. Therefore, orthogonality can also be used as a tool to establish (human) relevance of NAMs.

Here, we propose to apply orthogonal method for biological NAMs as a way to better characterise their relevance. For each of the screening NAMs different measuring principles exist. For instance, cytotoxicity can be assessed by simple live/ dead staining, by measuring release of lactate dehydrogenase (LDH) or by several other methods (see Usmani et al., 2024). The scientific plausibility of a given outcome can be improved when demonstrating that other methods (similar or orthogonal) deliver a comparable outcome.

## 3.3 Example: Triculture model of human small intestine to evaluate uptake and crossing of nanoparticles and barrier impairment

Hereunder an example is provided to showcase the application of the proposed qualification system. This specific NAM was selected as it addresses a key step in the application of the EFSA framework for NM risk assessment, i.e. the evaluation of any potential uptake of particles by the human small intestinal epithelium and of their translocation into the basolateral compartment, which implies that the translocated particles become systemically available. Also, the same *in vitro* model allows to gain evidence about potential barrier impairment resulting from interaction with the particles. Assessing intestinal particle uptake and crossing is extremely challenging *in vivo* and using a relevant model of the human small intestinal epithelium reduces the technical complexity of the assessment and increases its reliability by focusing on a human-relevant model specifically developed to take particle-specific uptake pathways into account. Therefore, this NAM provides an appropriate example of how conventional chemical risk assessment should be integrated with a nano-specific assessment addressing targeted nanoscale considerations.

In the following sections, the *in vitro* model is characterised according to the proposed method readiness criteria in line with the level of information foreseen to be provided at the phase of the submission of the LoI for a cell-based biological NAM.

### 3.3.1 General Information

**Name of the NAM:** Triculture model of human small intestine to evaluate uptake and crossing of nanoparticles (NPs) and barrier impairment.

**Synonyms:** intestinal epithelial Caco-2/HT29/Raji-B model; Caco-2/HT29/Raji-B co-culture (or 'triple culture').

**Summary:** This model represents a significant advancement in *in vitro* approaches to study the interaction of NPs with the human small intestinal epithelium compared to Caco-2 monocultures, the commonly used model to assess absorption of chemicals. Cells are cultured in a transwell-system characterised by an apical (AL) compartment, representing the intestine luminal surface exposed to the NM, and by a basolateral (BL) compartment, representing the underlying tissue. Enterocytes-like Caco-2 cells are plated together with mucus-producing HT29-MTX cells in the AL compartment. At 14 days of culture, Raji-B lymphocytes are added to the BL compartment to promote the differentiation of Caco-2 cells into microfold cells (M-cells). At 20 days of co-culture, the intestinal tight barrier is completely developed. Detailed SOPs are available. The presence in the model of mucus-secreting goblet cells and M-cells is essential as any particle reaching the intestinal epithelium has to cross the mucus layer first and because particle translocation may largely take place via M-cells. Whenever feasible, cell uptake is measured quantitatively, e.g. via mass spectrometry detection. Particle translocation is determined by analysing the BL medium with the same methodology used for measuring uptake, in case combined with ultrafiltration. Functional readouts include Transepithelial Electrical Resistance (TEER) and Lucifer Yellow (LY) translocation to characterise the proper differentiation of the system before exposure to particles and any impact of particle exposure on barrier function. The specific context-of-use is risk assessment

of NMs which do not readily dissolve during gastrointestinal digestion and reach the small intestine retaining their particulate nature.

**Intended use:** This NAM addresses a key step in nano-specific risk assessment, i.e. the evaluation of any potential uptake of particles by the human small intestinal epithelium and of their translocation into the BL compartment, which implies that the translocated particles become systemically available. Assessing intestinal particle uptake and crossing is extremely challenging *in vivo*. This NAM reduces the technical complexity of the assessment and increases its reliability by focusing on a human-relevant model specifically developed to take particle-specific uptake pathways into account. Also, the NAM provides evidence about potential barrier impairment resulting from interaction of the intestinal epithelium with the particles.

**Authors (for this example):** Olimpia Vincentini[1], Susanne Bremer-Hoffmann[2], Deborah Stanco[2], Alessia Bogni[2], Francesco Cubadda[1]

**Affiliations:**
[1] Istituto Superiore di Sanità - National Institute of Health (ISS), Rome, Italy;
[2] European Commission, Joint Research Centre (JRC), Ispra, Italy.

**Type of NAM:** Biological (low throughput *in vitro* cell model).

**Development Status:** RRL 6.[15]

**Transparency on proprietary elements:** There are no proprietary items to consider. According to the free patent database from the European Patent Office (Espacenet), patents related to the triculture model have not yet been issued.

**SOPs:** Different SOPs deriving from EU framework projects (e.g., *Patrols*, *NanoHarmony*) are available and were used as starting points in the EFSA-funded project NANOCELLUP (Vincentini et al., 2023). These SOPs have been developed for different types of nanomaterials including metal oxides and nanocellulose (the latter in NANOCELLUP). Key publications are Vincentini et al., 2022 and Vincentini et al., 2023. Updated SOPs for the assessment of metal oxide particles are being delivered within the NAMS4NANO project. The US National Institute of Standards and Technology (NIST) is currently adapting the protocol to nanoplastics. The current evaluation is based on the information gained as part of the NANOCELLUP project.

### 3.3.2 Detailed Description of the NAM

**Assay**

The test concentration should be selected considering the physicochemical properties of the material. E.g. for light, carbonaceous materials, lower mass concentrations must be applied as compared to inorganic NMs to expose cells to comparable particle number concentrations. The use of dispersion protocols ensuring maximal deagglomeration of particles is mandatory.

The intracellular uptake and translocation of NPs are measured, as a rule, after 24-hour exposure on the 21st day of culture (Vincentini et al., 2023, see Annex O). Barrier integrity

[15]Transitioning to RRL 7 (Guidance document under development at the OECD level).

and permeability can be concurrently measured. Shorter exposure time may be used if justified. Different exposure scenarios can also be considered, e.g. using a prolonged exposure time or repeated exposure conditions (Vincentini et al., 2023).

The intracellular uptake and translocation of NPs can be assessed qualitatively by using imaging techniques such as confocal laser scanning microscopy (CLSM), after fluorescent staining or labelling of the particles (see Vincentini et al., 2023; see Annex M), or electron microscopy, e.g. TEM. However, due to the technical challenges associated to the detection of particles at low levels in cells via EM, the absence of detection cannot represent a proof of absence of particles.

Quantitative assessment of intracellular uptake and translocation of nanoparticles is possible via chemical specific-methods. For instance, inductively coupled plasma mass spectrometry (ICP-MS) is applicable for inorganic nanoparticles, but it does not discriminate particulate and soluble forms. By adding an ultrafiltration step or resorting to single particle ICP-MS (spICP-MS), particles can be measured selectively. For assessing cell uptake, particle extraction is necessary and has to be performed without altering the particles present. For studying particle crossing, direct determination of particles in the BL medium is in many cases possible, for instance when spICP-MS is used. For particles with an organic chemical composition, analysis of the BL medium can be undertaken before and after ultrafiltration to discriminate particles and any potentially co-occurring dissolved counterpart.[16]

The disruption of the physiological barrier is considered as an important endpoint indicating a potential hazard of NMs. Barrier integrity can be assessed by measuring (TEER and LY translocation: these two endpoints allow to investigate the impact of ingested materials on the intestinal barrier function (Vincentini et al., 2023, see Annex O).

The model may be used for investigating possible local effects as secondary endpoints, e.g. cytotoxicity. For cytotoxicity, several well-known methods are suitable including the LDH, Alamar blue and MTT assays.

For all the above-mentioned measurements, no specialised equipment/custom-made instruments are needed.

**Biological system**

The system represents the human small intestinal epithelial tissue. It is a co-culture of three human cell lines. The physiological functions meant to be reproduced are the cell uptake and translocation processes of particles coming into contact with the human small intestine.

The three cell lines can be obtained by commercial suppliers.

The specific characteristics of the biological system are as follows. The triculture is obtained by culturing cells in a transwell-system characterised by an AL compartment, representing the luminal surface of the intestine exposed to the NM, and by a BL compartment, representing the underlying tissue (Figure 13). Enterocytes-like Caco-2 cells (Cellosaurus:

[16]The concentration (mass fraction) measured before ultrafiltration represents the sum of the particulate and dissolved fractions, whereas after ultrafiltration only the dissolved fraction (if present) is detected.

CVCL_002517; ATCC=HTB-3718) are plated in a 9:1 ratio together with the mucus producing HT29-MTX cells (Cellosaurus: CVCL_G35619) in the AL compartment. At 14 days of culture, Raji-B lymphocytes (Cellosaurus: CVCL_051120) are added to the BL compartment to promote the differentiation of Caco-2 cells into microfold cells (M-cells) (Vincentini et al., 2022). Finally, at 20 days of co-culture, the intestinal tight barrier is completely differentiated.

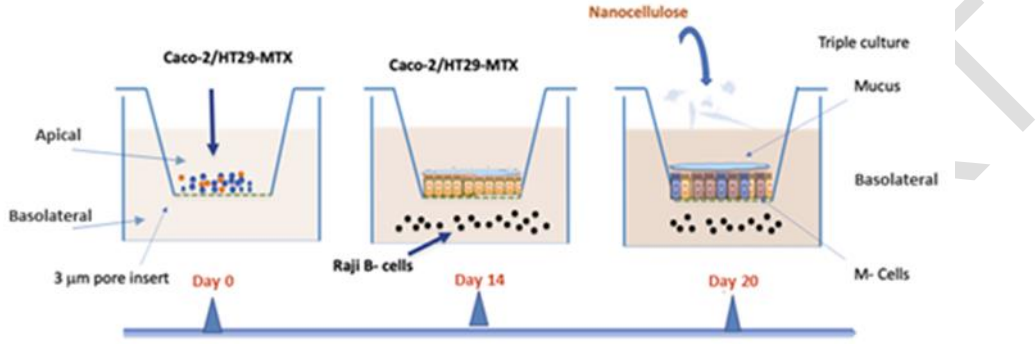


Figure 13: Scheme of the triculture model as applied in the NANOCELLUP project for assessing uptake and translocation of nanocellulose test materials

Detailed SOPs have been published for the cell culture and differentiation protocol. The most recent and comprehensive one is found in Vincentini et al., 2023 (see Annex O).

**Readout**

Particle uptake can be determined qualitatively (e.g. by fluorescence staining and use of CLSM) or quantitatively, for instance using labeling approaches or determining the characteristic element of an inorganic particle (e.g. Zn for ZnO) by ICP-MS and/or spICP-MS after particle extraction from cells. For particles with an organic chemical composition, methods such as HPLC-MS(/MS) can be used for proper chemical identification, particle translocation is determined by directly analysing the BL medium with the same methodology used for measuring uptake. Ultrafiltration should be applied to discriminate the particulate fraction from any potentially co-occurring dissolved fraction.

Functional readouts include TEER and LY translocation. They allow to characterise the proper development of the system before exposure to particles and any impact of particle exposure on barrier function. TEER provides functional information about barrier integrity. Its measurement is non-invasive and it is applied to monitor the complete barrier formation at the 20th day of differentiation (Srinivasan et al., 2015). At this time point, the triculture generates a lower TEER in comparison to the TEER of Caco-2 monocultures at the same timepoint.[17]

[17]Based on the currently existing dataset, the triculture generates a TEER with a minimum value of 200 $\Omega cm^2$ and with a fold change in the order of approximately -30% in comparison to the TEERs of Caco-2 monocultures at the

The paracellular LY passage, expressed as Apparent Permeability (Papp) values (cm/sec), is also evaluated as a quality criterion to assess the intestinal barrier permeability. The triculture shows a Papp increase in comparison to the value observed in the Caco-2 control at the 20th day of culture.[18]

### 3.3.3 Relevance for the context-of-use

The intestinal triculture model represents a significant advancement in *in vitro* approaches to study the interaction of NPs with the human intestinal epithelium (Lozoya-Agullo et al., 2017; Garcia-Rodriguez et al., 2018; Busch et al., 2021; Vincentini et al., 2022). This multiple cell culture allows to create a more representative and physiologically relevant model of the intestinal environment than the Caco-2 monoculture.

Caco-2 static monolayer cultures are a gold standard for *in vitro* assays of intestinal barrier properties accepted by regulatory authorities in ADME and toxicological studies of substances and may be useful to elucidate the absorption of such substances across the intestinal epithelium and their metabolism (Hu et al., 2004; Sun et al., 2008). Caco-2 cells are considered as the best intestinal epithelium model for chemicals taken up via the transcellular pathway, a most common drug permeation route (Dahlgren et al., 2019), although active passage as well as gut metabolism of drugs are thought to be drastically underestimated (Sarmento et al., 2012; Artursson et al., 2012). The absence of a physiologically relevant mucus layer and M-cells inherently limit the capability of this model of assessing uptake and crossing of NPs.

The triculture model mimics the human intestinal epithelium acting as barrier between the luminal environment and the bloodstream better than Caco-2 cell monocultures and thus best effectively predicts the potential of NMs to pass this barrier and become systemically available (Lozoya-Agullo et al., 2017; Garcia-Rodriguez et al., 2018; Busch et al., 2021; Vincentini et al., 2022). Whereas only very small particles can squeeze into intestinal cells (i.e. translocate via paracellular transport) (Ejazi et al., 2023), most particles are taken up by intestinal cells through transcellular transport via different mechanisms, directly linked to the particle physicochemical characteristics. In cases of gut inflammation or damage of the intestinal lining, the tight junctions between cells can become loose or disrupted. This creates larger gaps that NPs can more easily pass through, increasing the likelihood of entry into the bloodstream, and also potential eliciting chronic inflammatory responses.

The triculture model employs three different cell types to mimic the complex environment of the human gut. As already highlighted Caco-2 cells, although of colon origin, spontaneously differentiate in long-term culture in enterocytes, i.e., the most abundant cells in the intestinal lining, forming microvilli and tight junctions, and presenting hydrolysing enzymes and carrier-mediated transport systems responsible for nutrient absorption. However, since NPs cross the intestinal barrier of the small intestine mainly via M-cells, Caco-2 monocultures may underestimate translocation (Fröhlich, 2018). M-cells, localized in the follicle-associated

same timepoint. The latter thus serve as functional controls. TEER values not reflecting the above trend indicate that the triculture model is not functioning properly.

[18]Based on the currently available dataset, the $P_{app}$ fold change in comparison to the values observed in the Caco-2 control is generally in the range ca. 2-4

epithelium of Peyer's patches, represent approximately 1% of the cell lining of the intestine and their phenotype in human has largely remained uncharacterized due to the lack of specific markers. However, they play a crucial role in immune function by sampling antigens from the gut lumen and delivering them by transcytosis to antigen presenting cells and lymphocytes located in a pocket-like structure of the gut associated lymphoid tissue to induce immune responses. M-cells are distinguished from intestinal enterocytes by distinctive morphological features. At the apical surface, M-cells display a poorly organized brush border with short irregular microvilli, in contrast to the highly organised brush border of enterocytes (Vincentini et al., 2022). In the triculture model, the simultaneous presence of hematopoietic cells (Raji-B cells) is necessary to promote the Caco-2 conversion to the specialized M-cell phenotype.

The intestinal epithelium is covered with a layer of mucus continually secreted by the goblet cells, that protects the intestinal epithelium and impacts on NP mobility. Any particles reaching the intestinal epithelium have to cross the mucus layer first. Therefore, the presence of mucus-secreting cells is essential to investigate the uptake and translocation of particles. The mucus secreted by the intestinal goblet HT-29-MTX cells, which contains several mucin proteins, enhances the physiological relevance of the model and its suitability to investigate particle internalisation and crossing.

The NM physicochemical properties such as size, shape, surface charge, and coating all influence how the particles interact with the mucus layer and, if this is crossed, with the cell membrane, as well as which endocytic pathway is triggered (Des Rieux et al., 2013). Endocytosis encompasses several specific mechanisms including clathrin-mediated endocytosis (CME) which is considered one major route for NM uptake in intestinal cells (Zhao et al., 2011; Stern et al., 2012; Lundquist et al., 2016). CME involves protein structures called clathrin, forming coated pits on the cell membrane. These pits engulf the NPs, forming a vesicle that pinches off and is transported inside the cell. Another mechanism that may operate is caveolin-mediated endocytosis. Caveolae are small flask-shaped invaginations in the cell membrane enriched with caveolin protein which are thought to be involved in the uptake of specific types of NMs, depending on their size and surface properties. A third mechanism is clathrin- and caveolin-independent endocytosis. This is a less well-understood endocytic pathway and currently it is not clear to what extent the tri-culture model may feature this pathway for the internalisation of NPs. Other protein complexes and mechanisms for internalizing NPs into gut cells exist. Macropinocytosis is another relevant pathway, probably used by M-cells. The mechanism involves the formation of large, fluid-filled vesicles that engulf extracellular material. Internalised NPs may, e.g., cause the production of free radicals, leading to oxidative stress and damage to gut cells. Such a mechanism may result in a reduction of the TEER, indicating damage of the intestinal barrier. As already highlighted, damage to the gut barrier can result in more particles (as well as other harmful substances) to become systemically available (Ganguly et al., 2011; Huang et al., 2021; Zhao et al., 2023).

In addition to assessing the particles potential to be taken up, translocate to the BL compartment and leak into the bloodstream, the triculture model can be used to investigate potential local effects on the gut tissue.

The triculture model is a central component of the OECD Project 4.158 GD on an IATA for intestinal fate of orally ingested nanomaterials approved by the OECD Working Party for the Test Guideline Program (WNT) in 2022. The guidance document aims to establish a conceptual framework to assess the NMs behaviour in the different digestive compartments (mouth, stomach, small intestine) and their interactions with the intestinal mucosa.

The EFSA Guidance on Nano-RA highlights the *in vitro* assessment of particle uptake/crossing, cellular localization, accumulation as well as disruption of the physiological barrier as essential information requirements for NMs (EFSA Scientific Committee, 2021a). In this context, the triculture model is meant to be used as a key component of IATAs for the risk assessment of food-relevant NMs. The translocation of NMs into the circulatory system is an essential piece of information to decide further steps in the risk assessment of a NM.

# 4 Conclusions and Recommendations

We proposed a general framework for a NAMs qualification system to risk assessment in the food and feed sector, which initially shall be established for nanomaterials (NM) risk assessment. It can be regarded as a practical tool that allows to a) evaluate NAMs regarding their scientific validity for a specific context-of-use, b) support the optimisation of NAMs for regulatory application and c) ultimately facilitate their regulatory implementation for risk assessments in the food and feed sector. It describes a sequential framework from initial submission (LoI), to interim submission of a qualification plan (qualification advice) and the final submission of the full qualification proposal (qualification opinion). Importantly, the system also offers flexibility such that NAMs that are well enough developed can take a fast track to qualification (and skip the interim submission). Our system covers all phases of the process, namely i) the submission phase, ii) the evaluation phase and iii) the outcome phase. The specific aims are to provide a clear structure of the process and the individual steps as well as to provide guidance for the method developers and the evaluating experts by explaining i) how to structure the submissions, ii) which type of information and evidence is required, iii) what will be the criteria for the evaluation, iv) what is the expected duration of the process, v) what are the possible outcomes and vi) what are the consequences of a particular outcome/ decision.

We recommend to initially establish the system by focusing on NAMs for NMs aligned with the EFSA Guidance on Nano-RA (EFSA Scientific Committee, 2021a). As a first step, we propose three areas, covering NAMs for a) NM physicochemical characterisation, b) characterisation of NM in relevant biological fluids (e.g., *in chemico* NAMs to assess solubility, dissolution/degradation); c) toxicity screening (specifically involving intestinal uptake/ crossing, genotoxicity, cytotoxicity, reactivity/ oxidative stress, (pro-)inflammatory responses and barrier impairment). The criteria to evaluate method readiness have been structured into three overarching sections, namely i) detailled test method description (preferably as SOPs) covering three parts (set-up of the NAM, application and evaluation phase), ii) reliability and iii) relevance (in the context of the regulatory application). In addition, general information on the NAM shall be provided.

Importantly, we have already initially tested the evaluation criteria for one selected example, the triculture system, which can be applied to investigate NM uptake/ transport across a barrier and to evaluate NM effects on barrier integrity.

We recommend to firstly test the system to evaluate the readiness of selected NAMs in the context of risk assessment case studies, such as those currently ongoing within Lot 2 of the NAMs4NANO Project. This will allow for a critical discussion of the proposed system, the suggested process and the proposed criteria. Therefore, this report is an interim version, which should mainly serve as a basis for a broader discussion with EFSA, Member States and relevant stakeholders. The document is meant to be submitted to public consultation. All the above-named elements will help to fine-tune the approach and produce an updated proposal.

# References


Araujo F and Sarmento B, 2013. Towards the characterization of an in vitro triple co-culture intestine cell model for permeability studies. Int J Pharm, 458(1), 128-134. https://doi.org/10.1016/j.ijpharm.2013.10.003

Bal-Price A, Crofton KM, Sachana M, Shafer TJ, Behl M, Forsby A, Hargreaves A, Landesmann B, Lein PJ, Louisse J, Monnet-Tschudi F, Paini A, Rolaki A, Schrattenholz A, Sunol C, van Thriel C, Whelan M and Fritsche E, 2015. Putative adverse outcome pathways relevant to neurotoxicity. Crit Rev Toxicol, 45(1), 83-91. https://doi.org/10.3109/10408444.2014.981331

Bal-Price A, Hogberg HT, Crofton KM, Daneshian M, FitzGerald RE, Fritsche E, Heinonen T, Hougaard Bennekou S, Klima S, Piersma AH, Sachana M, Shafer TJ, Terron A, Monnet-Tschudi F, Viviani B, Waldmann T, Westerink RHS, Wilks MF, Witters H, Zurich MG and Leist M, 2018. Recommendation on test readiness criteria for new approach methods in toxicology: Exemplified for developmental neurotoxicity. ALTEX, 35(3), 306-352. https://doi.org/10.14573/altex.1712081

Balls M, Blaauboer BJ, Fentem JH, Bruner L, Combes RD, Ekwall B, Fielder RJ, Guillouzo A, Lewis RW, Lovell DP, Reinhardt CA, Repetto G, Sladowski D, Spielmann H and Zucco F, 1995. Practical Aspects of the Validation of Toxicity Test Procedures - the Report and Recommendations of Ecvam Workshop-5. Atla-Alternatives to Laboratory Animals, 23(1), 129-147. https://doi.org/10.1016/0887-2333(95)00082-8

Busch M, Kampfer AAM and Schins RPF, 2021. An inverted in vitro triple culture model of the healthy and inflamed intestine: Adverse effects of polyethylene particles. Chemosphere, 284, 131345. https://doi.org/10.1016/j.chemosphere.2021.131345

Crouzet T, Grignard E, Brion F, Blanc EB, Podechard N, Langouet S, Alonso-Magdalena P, Hubert P, Kim MJ and Audouze K, 2023. ReadEDTest: A tool to assess the readiness of in vitro test methods under development for identifying endocrine disruptors. Environ Int, 174, 107910. https://doi.org/10.1016/j.envint.2023.107910

Dahlgren D and Lennernas H, 2019. Intestinal Permeability and Drug Absorption: Predictive Experimental, Computational and In Vivo Approaches. Pharmaceutics, 11(8). https://doi.org/10.3390/pharmaceutics11080411

des Rieux A, Pourcelle V, Cani PD, Marchand-Brynaert J and Preat V, 2013. Targeted nanoparticles with novel non-peptidic ligands for oral delivery. Adv Drug Deliv Rev, 65(6), 833-844. https://doi.org/10.1016/j.addr.2013.01.002

European Commission, 2022. Commission Recommendation of 10.6.2022 on the definition of nanomaterial. Official Journal of the European Union. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022H0614(01)

EFSA (European Food Safety Authority), 2010. Application of systematic review methodology to food and feed safety assessments to support decision making. EFSA Journal 2010; 8(6): 1637. https://doi.org/10.2903/j.efsa.2010.1637

EFSA (European Food Safety Authority), 2014. Scientific opinion on good modelling practice in the context of mechanistic effect models for risk assessment of plant protection products EFSA Journal 2014, 12 (3). https://doi.org/10.2903/j.efsa.2014.3589

EFSA Scientific Committee, 2014. Modern methodologies and tools for human hazard assessment of chemicals. EFSA Journal, 2014; 12(4):3638, 87 pp. https://doi.org/10.2903/j.efsa.2014.3638

EFSA Scientific Committee, 2021a. Guidance on risk assessment of nanomaterials to be applied in the food and feed chain: human and animal health. EFSA Journal 2021, 19(8), e06768. https://doi.org/10.2903/j.efsa.2021.6768

EFSA Scientific Committee, 2021b. Guidance on technical requirements for regulated food and feed product applications to establish the presence of small particles including nanoparticles. EFSA Journal 2021, 19(8), e06769. https://doi.org/10.2903/j.efsa.2021.6769

EFSA (European Food Safety Authority), Tarazona J, Kass G, Dorne J-L, Liem D,Paraskevopoulos K, Kleiner J, Heppner C, Hugas M, 2022. New Approach Methodologies. EFSAsupporting publication 2022:e200502. 8 pp. doi:10.2903/sp.efsa.2022.e200502

Ejazi SA, Louisthelmy R and Maisel K, 2023. Mechanisms of Nanoparticle Transport across Intestinal Tissue: An Oral Delivery Perspective. ACS Nano, 17(14), 13044-13061. https://doi.org/10.1021/acsnano.3c02403

EMA (European Medicines agency), 2008. Qualification of novel methodologies for drug development: guidance to applicants. https://www.ema.europa.eu/en/qualification-novel-methodologies-medicine-development

Escher SE, Partosch F, Konzok S, Jennings P, Luijten M, Kienhuis A, de Leeuw V,Reuss R, Lindemann K-M, Hougaard Bennekou S, 2022. Development of a Roadmap for Action on NewApproach Methodologies in Risk Assessment. 153 pp. doi:10.2903/sp.efsa.2022.EN-7341

FDA (Food and Drug Administration), 2017. FDA's Predictive Toxicology Roadmap. Available online: https://www.fda.gov/media/109634/download

FDA (Food and Drug Administration), 2020. Qualification Process for Drug Development Tools – Guidance for Industry and FDA Staff. Available online: https://www.fda.gov/media/133511/download

FDA (Food and Drug Administration), 2023. Qualification Process for Medical Devices Development Tools – Guidance for Industry and FDA Staff. Available online: https://www.fda.gov/media/87134/download

Fröhlich E, 2018. Comparison of conventional and advanced in vitro models in the toxicity testing of nanoparticles. Artif Cells Nanomed Biotechnol, 46(sup2), 1091-1107. https://doi.org/10.1080/21691401.2018.1479709

Ganguly P, Breen A and Pillai SC, 2018. Toxicity of Nanomaterials: Exposure, Pathways, Assessment, and Recent Advances. ACS Biomater Sci Eng, 4(7), 2237-2275. https://doi.org/10.1021/acsbiomaterials.8b00068

Garcia-Rodriguez A, Vila L, Cortes C, Hernandez A and Marcos R, 2018. Exploring the usefulness of the complex in vitro intestinal epithelial model Caco-2/HT29/Raji-B in nanotoxicology. Food Chem Toxicol, 113, 162-170. https://doi.org/10.1016/j.fct.2018.01.042

Gourmelon A, Hubert P, Grignard E, Baumann L, Munn S and Michel C, 2024. The benefits of validation of methods for toxicity testing outweigh its costs. ALTEX. https://doi.org/10.14573/altex.2403051

Groenewold M, Bleeker EAJ, Noorlander CW, Sips A, van der Zee M, Aitken RJ, Baker JH, Bakker MI, Bouman EA, Doak SH, Drobne D, Dumit VI, Florin MV, Fransman W, Gonzalez MM, Heunisch E, Isigonis P, Jeliazkova N, Jensen KA, Kuhlbusch T, Lynch I, Morrison M, Porcari A, Rodríguez-Llopis I, Pozuelo BM, Resch S, Säämänen AJ, Serchi T, Soeteman-Hernandez LG, Willighagen E, Dusinska M and Scott-Fordsmand JJ, 2024. Governance of advanced materials: Shaping a safe and sustainable future. NanoImpact, 100513. https://doi.org/10.1016/j.impact.2024.100513

Hartung T, Bremer S, Casati S, Coecke S, Corvi R, Fortaner S, Gribaldo L, Halder M, Hoffmann S, Roi AJ, Prieto P, Sabbioni E, Scott L, Worth A and Zuang V, 2004. A modular approach to the ECVAM principles on test validity. Altern Lab Anim, 32(5), 467-472. https://doi.org/10.1177/026119290403200503

Huang X and Tang M, 2021. Review of gut nanotoxicology in mammals: Exposure, transformation, distribution and toxicity. Sci Total Environ, 773, 145078. https://doi.org/10.1016/j.scitotenv.2021.145078

ISO (International Organisation for standardisation), 2024. ISO/IEC GUIDE 98-1:2024. Guide to the expression of uncertainty in measurement (GUM)- Part 1: Introduction. Available online: https://www.vde-verlag.de/iec-normen/252561/iso-iec-guide-98-1-2024.html

JRC (Joint Research Centre), 2023. Zuang V, Daskalopoulos E, Berggren E, Batista Leite S, Bopp S, Carpi D, Casati S, Corvi R, Cusinato A, Deceuninck P, Dura A, Franco A, Gastaldello A, Gribaldo L, Holloway M, Katsanou E, Langezaal I, Morath S, Munn S, Prieto Peraita MDP, Piergiovanni M, Whelan M, Wittwehr C, Worth A and Viegas Barroso JF. Non-animal Methods in Science and Regulation, EUR 31395. Available online: https://publications.jrc.ec.europa.eu/repository/handle/JRC132525

Kestens V, Bozatzidis V, De Temmerman PJ, Ramaye Y and Roebben G, 2017. Validation of a particle tracking analysis method for the size determination of nano- and microparticles. J Nanopart Res, 19(8), 271. https://doi.org/10.1007/s11051-017-3966-8

Krebs A, Waldmann T, Wilks MF, Van Vugt-Lussenburg BMA, Van der Burg B, Terron A, Steger-Hartmann T, Ruegg J, Rovida C, Pedersen E, Pallocca G, Luijten M, Leite SB, Kustermann S, Kamp H, Hoeng J, Hewitt P, Herzler M, Hengstler JG, Heinonen T, Hartung T, Hardy B, Gantner F, Fritsche E, Fant K, Ezendam J, Exner T, Dunkern T, Dietrich DR, Coecke S, Busquet F, Braeuning A, Bondarenko O, Bennekou SH, Beilmann M and Leist M, 2019. Template for the description of cell-based toxicological test methods to allow evaluation and regulatory use of the data. ALTEX, 36(4), 682-699. https://doi.org/10.14573/altex.1909271

Linsinger TP, Chaudhry Q, Dehalu V, Delahaut P, Dudkiewicz A, Grombe R, von der Kammer F, Larsen EH, Legros S, Loeschner K, Peters R, Ramsch R, Roebben G, Tiede K and Weigel S, 2013. Validation of methods for the detection and quantification of engineered nanoparticles in food. Food Chem, 138(2-3), 1959-1966. https://doi.org/10.1016/j.foodchem.2012.11.074

Lozoya-Agullo I, Araujo F, Gonzalez-Alvarez I, Merino-Sanjuan M, Gonzalez-Alvarez M, Bermejo M and Sarmento B, 2017. Usefulness of Caco-2/HT29-MTX and Caco-2/HT29-MTX/Raji B Coculture Models To Predict Intestinal and Colonic Permeability Compared to Caco-2 Monoculture. Mol Pharm, 14(4), 1264-1270. https://doi.org/10.1021/acs.molpharmaceut.6b01165

Lundquist P and Artursson P, 2016. Oral absorption of peptides and nanoparticles across the human intestine: Opportunities, limitations and studies in human tissues. Adv Drug Deliv Rev, 106(Pt B), 256-276. https://doi.org/10.1016/j.addr.2016.07.007

OECD (Organization for Economic Cooperation and Development), 1998. OECD Principles on Good Laboratory Practice. Available online: https://doi.org/10.1787/9789264078536-en

OECD (Organization for Economic Cooperation and Development), 2005. Guidance document on the validation and international acceptance of new or updated test methods for hazard assessment. Series on Testing and Assessment No.34. Available online: https://one.oecd.org/document/ENV/JM/MONO(2005)14/en/pdf

OECD (Organization for Economic Cooperation and Development), 2010. Guidance document on using cytotoxicity tests to estimate starting doses for acute oral systematic toxicity

tests. OECD Series on Testing and Assessment, No. 129. Available online: https://one.oecd.org/document/env/jm/mono(2010)20/en/pdf

OECD (Organization for Economic Cooperation and Development), 2014. Guidance Document for Describing Non-Guideline In Vitro Test Methods. Series on Testing and Assessment No. 211. Available online: https://doi.org/10.1787/9789264274730-en

OECD (Organization for Economic Cooperation and Development), 2016a. Test No. 473: In Vitro Mammalian Chromosomal Aberration Test. OECD Guidelines for the Testing of Chemicals: Section 4. Available online: https://doi.org/10.1787/9789264264649-en

OECD (Organization for Economic Cooperation and Development), 2016b. Test No. 476: In Vitro Mammalian Cell Gene Mutation Tests using the Hprt and xprt genes. OECD Guidelines for the Testing of Chemicals: Section 4. Available online: https://doi.org/10.1787/9789264264809-en

OECD (Organization for Economic Cooperation and Development), 2016c. Test No. 490: In Vitro Mammalian Cell Gene Mutation Tests Using the Thymidine Kinase Gene. OECD Guidelines for the Testing of Chemicals: Section 4. Available online: https://doi.org/10.1787/9789264264908-en

OECD (Organization for Economic Cooperation and Development), 2018. Guidance Document on Good In Vitro Method Practices (GIVIMP). Series on Testing and Assessment, No. 286. Available online: https://doi.org/10.1787/9789264304796-en

OECD (Organization for Economic Cooperation and Development), 2019a. Test No. 431: In vitro skin corrosion: reconstructed human epidermis (RHE) test method. OECD Guidelines for the Testing of Chemicals, Section 4. Available online: https://doi.org/10.1787/9789264264618-en

OECD (Organization for Economic Cooperation and Development), 2019b. Test No. 432 In Vitro 3T3 NRU Phototoxicity Test. Available online: https://doi.org/10.1787/9789264071162-en

OECD (Organization for Economic Cooperation and Development), 2019c. Test No. 495: Ros (Reactive Oxygen Species) Assay for Photoreactivity. OECD Guidelines for the Testing of Chemicals: Section 4. Available online: https://doi.org/10.1787/915e00ac-en

OECD (Organization for Economic Cooperation and Development), 2020. Test No. 471: Bacterial Reverse Mutation Test. Available online: https://doi.org/10.1787/9789264071247-en

OECD (Organization for Economic Cooperation and Development), 2021a. Test No. 439: In Vitro Skin Irritation: Reconstructed Human Epidermis Test Method. Available online: https://doi.org/10.1787/9789264242845-en

OECD (Organization for Economic Cooperation and Development), 2021b. Guidance document on the characterisation, validation and reporting of Physiologically Based Kinetic (PBK) models for regulatory purposes. Series on Testing and Assessment, No. 331. Available online: https://doi.org/10.1787/d0de241f-en

OECD (Organization for Economic Cooperation and Development), 2022. Study Report on Applicability of the key event-based TG 442D for in vitro skin sensitisation testing of nano-materials. Available online: https://www.oecd.org/chemicalsafety/testing/draft-study-report-test-guideline-442D-in-vitro-skin-sensitisation-nanomaterials.pdf

OECD (Organization for Economic Cooperation and Development), 2023a. Test No. 487: In Vitro Mammalian Cell Micronucleus Test. OECD Guidelines for the Testing of Chemicals: Section 4. Available online: https://doi.org/10.1787/9789264264861-en

OECD (Organization for Economic Cooperation and Development), 2023b. Test No. 498: In vitro Phototoxicity - Reconstructed Human Epidermis Phototoxicity test method. OECD Guidelines for the Testing of Chemicals: Section 4. Available online: https://doi.org/10.1787/7b2f9ea0-en

OECD (Organization for Economic Cooperation and Development), 2023c. (Q)SAR Assessment Framework: Guidance for the regulatory assessment of (Quantitative) Structure Activity Relationship models and predictions. Series on Testing and Assessment, No. 386. Available online: https://doi.org/10.1787/d96118f6-en

OECD (Organization for Economic Cooperation and Development), 2023d. Test No. 126: Determination of the Hydrophobicity Index of Nanomaterials Through an Affinity Measurement. OECD Guidelines for the Testing of Chemicals: Section 1. Physical-Chemical properties. Available online: https://doi.org/10.1787/ae9c0fd1-en

OECD (Organization for Economic Cooperation and Development), 2023e. Test No. 492: Reconstructed human Cornea-like Epithelium (RhCE) test method for identifying chemicals not requiring classification and labelling for eye irritation or serious eye damage. OECD Guidelines for the Testing of Chemicals: Section 4. Available online: https://doi.org/10.1787/9789264242548-en

OECD (Organization for Economic Cooperation and Development), 2023f. Test Guideline No. 442C: In Chemico Skin Sensitisation: Assays addressing the Adverse Outcome Pathway key event on covalent binding to proteins. Available online: https://doi.org/10.1787/9789264229709-en

OECD (Organization for Economic Cooperation and Development), 2023g. Guideline No. 497: Defined Approaches on Skin Sensitisation. OECD Guidelines for the Testing of Chemicals: Section 4. Available online: https://doi.org/10.1787/b92879a4-en

OECD (Organization for Economic Cooperation and Development), 2023h. Report on the WNT Workshop how to prepare the Test Guidelines Programme for emerging technologies. Series on Testing and Assessment, No. 378. Available online: https://one.oecd.org/document/ENV/CBC/MONO(2023)14/en/pdf

OECD (Organization for Economic Cooperation and Development), 2023i. Test No. 125: Nanomaterial Particle Size and Size Distribution of Nanomaterials. OECD Guidelines for the Testing of Chemicals, Section 1: Physical-Chemical properties. Available online: https://doi.org/10.1787/af5f9bda-en

OECD (Organization for Economic Cooperation and Development), 2023j. Test No. 437: Bovine Corneal Opacity and Permeability Test Method for Identifying i) Chemicals Inducing Serious Eye Damage and ii) Chemicals Not Requiring Classification for Eye Irritation or Serious Eye Damage. Available online: https://doi.org/10.1787/9789264203846-en

OECD (Organization for Economic Cooperation and Development), 2024. Decision of the Council concerning OECD Legal Instruments the Mutual Acceptance of Data in the Assessment of Chemicals. OECD Legal Instruments. Available online: https://www.oecd.org/chemicalsafety/testing/council-acts-on-mutual-acceptance-of-data.htm

Sarmento B, Andrade F, da Silva SB, Rodrigues F, das Neves J and Ferreira D, 2012. Cell-based in vitro models for predicting drug permeability. Expert Opin Drug Metab Toxicol, 8(5), 607-621. https://doi.org/10.1517/17425255.2012.673586

Schmeisser S, Miccoli A, von Bergen M, Berggren E, Braeuning A, Busch W, Desaintes C, Gourmelon A, Grafstrom R, Harrill J, Hartung T, Herzler M, Kass GEN, Kleinstreuer N, Leist M, Luijten M, Marx-Stoelting P, Poetz O, van Ravenzwaay B, Roggeband R, Rogiers V, Roth A, Sanders P, Thomas RS, Marie Vinggaard A, Vinken M, van de Water B, Luch A and Tralau

T, 2023. New approach methodologies in human regulatory toxicology - Not if, but how and when! Environ Int, 178:108082. doi: 10.1016/j.envint.2023.108082

Shandilya N, Barreau MS, Suarez-Merino B, Porcari A, Pimponi D, Jensen KA, Fransman W and Franken R, 2023. TRAAC framework to improve regulatory acceptance and wider usability of tools and methods for safe innovation and sustainability of manufactured nanomaterials. NanoImpact, 30, 100461. https://doi.org/10.1016/j.impact.2023.100461

Simon CG Jr, Borgos SE, Calzolai L, Nelson BC, Parot J, Petersen EJ, Roesslein M, Xu X and Caputo F, 2023. Orthogonal and complementary measurements of properties of drug products containing nanomaterials. J Control Release, 354, 120-127. https://doi.org/10.1016/j.jconrel.2022.12.049

Srinivasan B, Kolli AR, Esch MB, Abaci HE, Shuler ML and Hickman JJ, 2015. TEER measurement techniques for in vitro barrier model systems. J Lab Autom, 20(2), 107-126. https://doi.org/10.1177/2211068214561025

Stern ST, Adiseshaiah PP and Crist RM, 2012. Autophagy and lysosomal dysfunction as emerging mechanisms of nanomaterial toxicity. Part Fibre Toxicol, 9, 20. https://doi.org/10.1186/1743-8977-9-20

Sun H, Chow EC, Liu S, Du Y and Pang KS, 2008. The Caco-2 cell monolayer: usefulness and limitations. Expert Opin Drug Metab Toxicol, 4(4), 395-411. https://doi.org/10.1517/17425255.4.4.395

Usmani SM, Bremer-Hoffmann S, Cheyns K, Cubadda F, Dumit VI, Escher SE, Fessard V, Gutleb AC, Léger T, Liu Y-C, Mast J, McVey E, Mertens B, Montalvo D, Oomen AG, Ritz V, Serchi T, Sieg H, Siewert K, Stanco D, Verleysen E, Vincentini O, Yeo CWS, Yu D, van der Zande M and Haase A, 2024. Review of New Approach Methodologies for Application in Risk Assessment of Nanoparticles in the Food and Feed Sector: Status and Challenges. EFSA supporting publication 2024:EN-8826. 96 pp. doi:10.2903/sp.efsa.2024.EN-8826

Verleysen E, Wagner T, Lipinski HG, Kagi R, Koeber R, Boix-Sanfeliu A, De Temmerman PJ and Mast J, 2019. Evaluation of a TEM based Approach for Size Measurement of Particulate (Nano)materials. Materials (Basel), 12. doi: 10.3390/ma12142274

Vincentini O, Prota V, Cecchetti S, Bertuccini L, Tinari A, Iosi F and De Angelis I, 2022. Towards the Standardization of Intestinal In Vitro Advanced Barrier Model for Nanoparticles Uptake and Crossing: The SiO(2) Case Study. Cells, 11(21), 3357. https://doi.org/10.3390/cells11213357

Vincentini O, Blier AL, Bogni A, Brun M, Cecchetti S, De Battistis F, Denis S, Etienne-Mesmin L, Ferraris F, Fumagalli FS, Hogeveen K, Iacoponi F, Raggi A, Siciliani L, Stanco D, Verleysen E, Fessard V, Mast J, Blanquet-Diot S, Bremer-Hoffmann S and Cubadda F, 2023. EFSA Project on the use of New Approach Methodologies (NAMs) for the hazard assessment of nanofibres. Lot 1, nanocellulose oral exposure: gastrointestinal digestion, nanofibres uptake and local effects. EFSA Supporting Publications, 20(9), 8258E. https://doi.org/10.2903/sp.efsa.2023.EN-8258

Westmoreland C, Bender HJ, Doe JE, Jacobs MN, Kass GEN, Madia F, Mahony C, Manou I, Maxwell G, Prieto P, Roggeband R, Sobanski T, Schutte K, Worth AP, Zvonar Z and Cronin MTD, 2022. Use of New Approach Methodologies (NAMs) in regulatory decisions for chemical safety: Report from an EPAA Deep Dive Workshop. Regul Toxicol Pharmacol, 135, 105261. https://doi.org/10.1016/j.yrtph.2022.105261

WHO (World Health Organisation), 2009 IPCS No.7. Assessment of combined exposures to multiple chemicals: report of a WHO/IPCS international workshop on aggregate/cumulative risk assessment. Available online: https://iris.who.int/handle/10665/44113

Zhao F, Zhao Y, Liu Y, Chang X, Chen C and Zhao Y, 2011. Cellular uptake, intracellular trafficking, and cytotoxicity of nanomaterials. Small, 7(10), 1322-1337. https://doi.org/10.1002/smll.201100001

Zhao Y, Liu S and Xu H, 2023. Effects of microplastic and engineered nanomaterials on inflammatory bowel disease: A review. Chemosphere, 326, 138486. https://doi.org/10.1016/j.chemosphere.2023.138486

# Glossary

Most terms are aligned with OECD GD 34 (OECD, 2005) and OECD GIVIMP (OECD, 2018).

Accuracy: The closeness of agreement between test method results and accepted reference values. For analytical methods this is understood as a property of a single measurement, which includes precision and trueness (expressed as bias) while for biological NAMs it is often used interchangeably with "concordance" to mean the proportion of correct outcomes of a test method (also see test method accuracy).

Adverse effect: A change in the morphology, physiology, growth, development, reproduction or life span of an organism, system or (sub)population that results in an impairment of functional capacity, an impairment of the capacity to compensate for additional stress, or an increase in susceptibility to other influences (WHO/IPCS, 2009).

Adverse Outcome Pathway (AOP): An AOP is an analytical construct that describes a sequential chain of causally linked events at different levels of biological organisation that lead to an adverse health or ecotoxicological effect.

Analytical NAM: Analytical NAMs describe analytical measurement procedures such as those commonly applied for NM physicochemical characterisation or those utilised to quantify migration/release. They may be seen as part of the wider group of *in chemico* NAMs.

Applicability domain: defines the (chemical) space within which a specific test method delivers meaningful results. Thus, it specifies for which chemicals (with which (physicochemical) properties) a test method is applicable.

Assay: Used interchangeably with test.

Biomarker: A biological parameter that is objectively measured and evaluated as an indicator of normal biological state or pathological processes.

Biological NAM: NAM that requires a biological model (*in vitro*, *ex vivo* cell models or whole organisms *in vitro* models) in order to investigate a response of that model upon a chemical treatment. More commonly such NAMs are referred to as *in vitro* tests.

Biological Model: Biological models of different complexity exist. Common examples are established cell lines, which can be obtained from a repository or primary cells that are directly taken from a living organism. Biological models can be composed of one cell type only or employ different cell types in co-culture and/or complex 3D architecture (e.g., 3D tissue models, organoids). Stem cell-derived models are also increasingly being exploited. More recently, there also is growing attention for whole organism models that involve simpler organisms such as *C. elegans*, which would not be classified as animal tests under the European legislation. The biological model is related to the test system.

Biological plausibility: The biological plausibility relies on an understanding of the fundamental biological processes involved and whether they are consistent with the causal relationship being proposed. In addition, in the context of the MoA/AOP frameworks, biological plausibility

is one of the elements to be considered in the weight of evidence analysis, where reference is made to the biological plausibility of the KE relationships.

Context-of-Use: refers to a clearly articulated description delineating the manner and purpose of use for the tool (when and how it will be used) as defined in the Predictive Toxicology Roadmap (FDA, 2017). In our case, the NAMs are intended to be used for NM risk assessments in the food and feed sector, specifically for the initial steps described in the EFSA Guidance on risk assessment of NMs (EFSA Scientific Committee, 2021a). They are intended to be applied for initial screening and NAM-based data shall be evaluated in an integrated manner together with other existing evidence. Thus, they are not intended as stand-alone methods for the prediction of an adverse effect.

Data: typically refers to "raw data", meaning the "raw form of an information" as directly obtained from a measurement.

Data analysis: means all processes that bring order, structure and meaning into data. This may involve different mathematical or statistical approaches that can be applied to data. A common example is a linear trend analysis.

Data evaluation: includes all the steps and procedures that are done with data, including "preprocessing", "analysis" and "interpretation". Because data evaluation needs considerable resources it is usually combined with data management.

Data interpretation: Once data analysis is completed an interpretation is needed to assign a "meaning", which is required in order to make decisions based on data. For instance, to decide when a test result is considered "positive" and when "negative" (possible examples are to compare the outcome of a treated sample with an untreated control and expecting a x-fold increase/decrease, with/without considering uncertainty). Data interpretation can be classified as quantitative or qualitative.

Data management: Proper data management is indispensable for every laboratory and includes considerations on data storage for raw, analysed and evaluated data (how to store such data internally within a reasonable time frame), data archiving (what to do with older data) and also on data publishing/ releasing (e.g. in a public database).

Data preprocessing: includes all preparatory steps that are required before a meaningful analysis can be initiated. Common examples are background substraction, normalisation, the identification/ removal of outliers or any transformation steps such as log-transformation of data.

Dissolution: The process by which a NM dissolves in a liquid medium into its respective ions or molecules, resulting in the loss of its nano-specific features. Dissolution is a kinetic process.

Dissolution rate: Dissolution is typically expressed as a dissolution rate, which is the amount of substance dissolved (solute) into a solvent over a certain time period.

Dose-response analysis: describes the change (in nature, incidence, magnitude and/or severity) in an effect on an organism caused by different levels of exposure (or doses) to a stressor (usually a chemical) after certain exposure duration. This definition includes the

following assumptions: the response observed is due to an agent administered, the magnitude of the response is in fact related to the dose and the observed effect is quantifiable. Dose-response analysis is part of the hazard assessment during the process of risk assessment usually with the aim to derive a threshold for toxicity.

Endocrine activity: Interaction with the endocrine system that can potentially result in a response of the endocrine system, target organs and tissues. A substance with an endocrine activity has the potential to alter the function(s) of the endocrine system.

Endocrine disruptor: An exogenous substance or mixture that alters function(s) of the endocrine system and consequently causes adverse health effects in an intact organism, or its progeny, or (sub)populations (WHO/IPCS, 2009).

Endpoint (of toxicity): Generally speaking, it is a description of what is the "observable outcome" of a specific assay in relation to a specific toxicity. The most obvious endpoint is "lethality" in relation to "acute toxicity". Other endpoints can be more diverse as they may relate to specific biochemical, physiological or behavioural changes observable in biological models.

Engineered nanomaterial: The term "engineered nanomaterial" is defined in the EU Regulation on Novel Foods (EU) 2015/2283 as "any intentionally produced material that has one or more dimensions of the order of 100 nm or less or that is composed of discrete functional parts, either internally or at the surface, many of which have one or more dimensions of the order of 100 nm or less, including structures, agglomerates or aggregates, which may have a size above the order of 100 nm but retain properties that are characteristic of the nanoscale." Nanoscale refers to a size range between 1 to 100 nm.

Exposure assessment: Quantitative assessment of the exposure of an individual or a (sub)population to an agent. Exposure assessment is an important step in the risk assessment process.

False negative: A substance incorrectly identified as negative or non-active by a test method, when in fact it is positive or active.

False negative rate: The proportion of all positive substances that are falsely identified by a test method as negative. The false negative rate is one indicator of the test method performance.

False positive: A substance incorrectly identified as positive or active by a test, when in fact it is negative or non-active.

False positive rate: The proportion of all negative (non-active) substances that are falsely identified as positive or active. The false positive rate is one indicator of the test performance.

Goodness-of-fit: Describes how well predicted values (as obtained from calculations) fit to experimentally obtained values (common term for *in silico* methods).

Harmonised reporting templates: Templates for reporting results considering data and metadata, which are agreed on (harmonised). The most important examples for chemical risk assessment are the OECD Harmonised Templates (OHTs).

Human relevance: The extent to which certain results can be applied to humans for a given purpose.

*In chemico* approach: means an experimental approach that is performed without the use of a biological model, i.e. abiotic approaches. Common examples are dissolution/degradation rate tests, reactivity assays (e.g. NM surface reactivity) or binding assays (e.g. receptor-ligand, peptide-chemical).

*In silico* approach: means an approach that uses computer-based estimations or simulations. Common examples include (quantitative) structure-activity relationships ((Q)SARs) or physiology-based kinetic (PBK) models.

*In vitro* approach: means an approach involving an *in vitro* model and investigating its response(s) towards a treatment (synonym: biological NAM).

Inter-laboratory reproducibility: A measure of the extent to which different qualified laboratories, using the same protocol and testing the same substances, can produce qualitatively and quantitatively similar results. Also referred to as between-laboratory reproducibility. Part of the reliability assessment of a method.

Intermediate Precision: refers to the agreement of results from within-laboratory variations due to random events that might occur when using the method on different days, involving different analysts, using different equipment and/or under different calibration. At least the variation of one of these factors has to be investigated. These studies are carried out according to ISO 5725-3 (2023).

Intra-laboratory reproducibility: A determination of the extent that the same laboratory can successfully replicate results using a specific protocol (e.g. at different times or by different users). Also referred to as within-laboratory reproducibility. Part of the reliability assessment of a method.

Key event: A change in biological or physiological state that is both measurable and essential to the progression of a defined biological perturbation leading to a specific adverse outcome.

Linearity: the ability of a method to obtain test results proportional to the value of the measurement, which needs to be demonstrated by a calibration over the intended working range of the method by using appropriate concentration intervals (common term for analytical methods).

Limit of Detection (LoD)/ Limit of Quantification (LoQ): this refers to the lowest concentration that can be reliably detected (LoD)/ reliably quantified (LoQ) (common term for quantitative analytical methods).

Mechanism of action: A detailed molecular description of the mechanistic interaction through which a substance/molecule produces its effect.

Mode of action (MoA): A biologically plausible sequence of key events at different levels of biological organisation, starting with the exposure to a chemical and leading to an observed (adverse) effect.

Metadata: are structured data that provide descriptive information on data. For instance, they describe important aspects on how the data was obtained (for example which SOP was applied), on specific test conditions (for example if serum was present during the measurement, the passage number of the cells) or on other important aspects, which are critical to understand the data.

Nanomaterial (NM): According to the updated Recommendation of the European Commission (2022/C 229/01), a NM is present if at least 50% of the number of solid particles it contains meet at least one of the following conditions:

- one or more external dimensions of the particles in the size range 1 nm to 100 nm;
- the particles have an elongated shape, with two external dimensions being less than 1 nm and the other external dimension being greater than 100 nm;
- the particles have a plate-like shape, with one external dimension being less than 1 nm and the other external dimensions being greater than 100 nm.

In the context of EFSA guidance documents on risk assessment of NM, the term is also applicable to the fraction of nanoparticles, present in conventional materials and any other material or fraction exhibiting characteristics of the nanoscale and, consequently, requiring risk assessment with specific considerations for addressing these characteristics.

New approach methodology (NAM): Any technology, methodology, approach that can provide useful information to support chemical risk assessment by informing on the hazard (considering toxicodynamics and toxicokinetics) and exposure of a chemical without the use of animal tests shall be considered a New Approach Method (NAM). This includes *in silico*, *in chemico*, *in vitro*, *ex vivo* and, specifically for NMs, also physicochemical characterisation approaches. Any non-animal method can qualify as a NAM when it is validated or when is has been demonstrated to be scientifically valid for the application in risk assessment (according to Usmani et al, 2024).

Optimised method: describes a method for which the test protocol(s) have been optimised, typically in iterative cycles. As a result, these protocols are available as SOPs. Test method optimisation is part of prevalidation.

Optimised test protocol: A test protocol that has been revised and improved based on the results obtained in prevalidation and validation studies.

Orthogonal method: Orthogonal methods use different physical principles to measure the same property of a sample. For NMs, particle size distribution can be measured by nanoparticle tracking analysis (NTA) or electron microscopy (EM), which are orthogonal methods as they use different physical principles. However, NTA and dynamic light scattering (DLS) are not orthogonal as they rely on the same physical principle (namely Brownian motion).

Performance criteria: The operational characteristics of a test are the measures of its performance under specific conditions, including its reliability and accuracy that describe the test's usefulness, limitations, and relevance.

Performance standards: Standards, based on a validated test method, that provide a basis for evaluating the comparability of a proposed test method that is mechanistically and functionally similar. Included are (1) essential test method components; (2) a minimum list of reference chemicals selected from among the chemicals used to demonstrate the acceptable performance of the validated test method; and (3) the comparable levels of accuracy and reliability, based on what was obtained for the validated test method, that the proposed test method should demonstrate when evaluated using the minimum list of reference chemicals.

Precision: Defined as the closeness of agreement among individual test results from repeated analyses of a homogeneous sample. It can be determined at three levels: repeatability, intermediate precision and reproducibility (common term for analytical methods).

Prevalidation: Multi-center validation studies require huge amounts of resources such that they require a good preparation, which often is referred to as prevalidation. However, there is no formal definition of this term. Often, the term is used synonymously with optimisation. To demonstrate the transferability of a method is an important part of prevalidation. In some cases, also initial (smaller) ILCs are conducted in this phase.

Predictive capacity: Established by comparing the results of a NAM with reference values (which are then taken as the "true values"), often such reference values are derived from *in vivo* studies or from human evidence (for biological NAMs).

Prediction model (PM): The methodology which is used to predict the degree of toxicity (for biological NAMs). More generally speaking this relates to the data interpretation procedure.

Predictivity (negative): The proportion of correct negative responses among substances testing negative by a test method. It is one indicator of test method accuracy. Negative predictivity is a function of the sensitivity of the test method and the prevalence of negatives among the substances tested.

Predictivity (positive): The proportion of correct positive responses among substances testing positive by a test method. It is one indicator of test method accuracy. Positive predictivity is a function of the sensitivity of the test method and the prevalence of positives among the substances tested.

Proprietary item/ element: refers to anything that is associated to a test method, where the use is restricted by patents, copyrights, trademarks, etc.

Protocol: The detailed, unambiguous step-by-step description that provides information on how to perform the test method (including preparations, reagents, supplies, and equipment needed, and also details on data evaluation). During the optimisation phase, the protocols are developed into SOPs.

(Q)SARs (Quantitative Structure-Activity Relationships): Theoretical models for making predictions of physicochemical properties, environmental fate parameters, or biological effects (including toxic effects in environmental and mammalian species). They can be divided into two major types, QSARs and SARs. QSARs are quantitative models yielding a continuous or categorical result while SARs are qualitative relationships in the form of structural alerts that incorporate molecular substructures or fragments related to the presence or absence of activity.

Qualification: is a conclusion that the results of an assessment using the model or assay can be relied on to have a specific interpretation and application in product development and regulatory decision-making within the stated context-of-use (Predictive Toxicology Roadmap, FDA, 2017). Thus, qualification is an expert opinion that facilitates the regulatory implementation of a NAM in one sector.

Reference chemicals: In the context of validation (OECD GD 34) these are chemicals, which were selected for use in the validation process. For those chemicals, the responses (in the *in vitro* or *in vivo* reference test system or the species of interest) are already known. These chemicals should be representative of the classes of chemicals for which the test method is expected to be used, and should represent the full range of responses that may be expected from the chemicals for which it may be used, from strong, to weak, to negative. Different sets of reference chemicals may be required for the different stages of the validation process, and for different test methods and test uses.

Reference data: An agreed-upon set of values against which the values obtained using the new test will be compared.

Reference test method: A test method against which the results from the new test method are being compared.

Regulatory acceptance: The formal acceptance of a test method by regulatory authorities indicating that the test method may be used to provide information to meet a specific regulatory requirement.

Relevance: Refers to the appropriateness of the data for the intended purpose of the assessment. This also includes the description of the predictive capacity and the applicability domain.

Reliability: Is commonly assessed by determining the intra- and interlaboratory reproducibility. Sometimes this is also referred to as robustness or repeatability.

Repeatability: Refers to the ability of the method to generate the same results over a short time interval under identical conditions. It is recommended to measure at least ten subsamples in one run or three subsamples over five different days (common term for analytical methods).

Reproducibility: see Reliability

Risk: The probability or degree of concern that a defined exposure to a substance will cause an adverse effect in the species of interest or in the environment. Risk is determined by combining information on hazard and exposure.

Risk assessment: A process intended to calculate or estimate the risk to a given target organism, system or (sub)population, including the identification of attendant uncertainties, following exposure to a particular agent, taking into account the inherent characteristics of the agent of concern as well as the characteristics of the specific target system. The Risk assessment process includes four steps: hazard identification, hazard characterisation (related term: dose-response assessment), exposure assessment, and risk characterisation.

Robust(ness): Is defined as a measure of its capacity to obtain comparable and acceptable results when perturbed by small but deliberate variations in procedural parameters listed in the documentation. (see Reliability)

SAR: See (Q)SARs.

Scientifically plausible method: In the context of the here proposed regulatory readiness levels (RRLs) the research phase of test method development ends with a "scientifically plausible method", which means that the underlying scientific principles of the method are clear (which is linked to biological plausibility for biological NAMs), an initial method description is available (which is not yet optimised) and the method has already been applied (for instance in scientific publications).

Screening test: means a rapid and simple test. The results are typically used for categorising substances for initial decision making (screening & prioritisation). In tiered testing strategies, screening tests are included in the first tier with the aim to reduce the number of substances that would require a more detailed, resource-intense follow-up in higher tiers. Typical screening tests are *in silico* models and (high throughput) *in vitro* tests that assess important key events such as cytotoxicity, oxidative stress or pro-inflammatory responses.

Selectivity: this is the ability of a method to correctly determine the analyte without interference from other compounds (common term for analytical methods).

Sensitivity: The proportion of all positive/active substances that are correctly classified by the test. It is a measure of accuracy for a test method that produces categorical results, and is an important consideration in assessing the relevance of a test method.

Solubility: The maximum amount of a solute that can dissolve in a given quantity of solvent at a specified temperature and pressure to form a saturated solution. Solubility is often expressed as the mass of solute per volume (g/L) or mass of solute per mass of solvent (g/g), or as the moles of solute per volume (mol/L). Solubility in relevant media requires description of the media and the conditions under which the measurements were made Solubility is a thermodynamic property while dissolution describes a kinetic process.

Specificity: The proportion of all negative/inactive substances that are correctly classified by the test. It is a measure of accuracy for a test method that produces categorical results and is an important consideration in assessing the relevance of a test method.

Standard Operating Procedure (SOP): here we refer to OECD GIVIMP (2018), where SOPs are understood as advanced (optimised) concise descriptions involving all details/ steps of a method. Ideally, a method is firstly optimised, during the optimisation a SOP shall be developed before afterwards a (in-house) validation is initiated. Importantly, GIVIMP stresses

that typically one method requires a set of SOPs (one SOP per activity, e.g. routine cell culture, maintenance/ calibration of devices, preparation of test items, conducting a specific measurement, data storage, data analysis etc). The reason why not to include all steps into one single document is that such a modular approach a) allows for easier use by the personnel, b) is easier to keep up to date and c) allows for greater flexibility (new approaches can build upon existing modular SOPs). SOPs always should be written in an active voice (first do this, then do that). It is advisable that they are written by the laboratory staff that is actually performing these activities on a daily basis, but each SOP shall be reviewed and approved, most commonly by the facility manager. Please note that SOPs are also requested within OECD GLP (1998). It is advisable to categorise SOPs by type (some are administrative, others are technical). Please note that in particular in nanosafety research several SOPs have been harmonised in research consortia, meaning that several experts of different institutes were involved in the optimisation and harmonisation. In several cases, these SOPs were also tested in ILCs.

Test (or assay): An experimental system used to obtain information on the adverse effects of a substance. Used interchangeably with assay.

Test battery: A series of tests usually performed at the same time or in close sequence. Each test within the battery is designed to complement the other tests and generally to measure a different component of a multi-factorial toxic effect or to confirm another test. Also called base set or minimum data set in ecotoxicological testing.

Test method: A process or procedure used to obtain information on the characteristics of a substance or agent. Toxicological test methods generate information regarding the ability of a substance or agent to produce a specified biological effect under specified conditions. Used interchangeably with "test" and "assay".

Test method accuracy: The closeness of agreement between test method results and accepted reference values. For analytical methods this is understood as a property of a single measurement, which includes precision and trueness (expressed as bias) while for biological NAMs it is often used interchangeably with "concordance" to mean the proportion of correct outcomes of a test method (also see test method accuracy).

Test method development: Within this document the term "test method development" shall be understood in a broader manner to cover all the phases until the method is "ready" for regulatory application. Hence, four major phases can be distinguished: i) research, ii) optimisation, iii) validation and iv) international harmonisation.

Test method submitter: An entity that submits a method for evaluation, which could be the developer of a method (alone or in a consortium), a business operation (who is interested to apply such a method) or a sponsor (who takes responsibility to organise and provide all the available, relevant information).

Test system: Any biological, chemical or physical system or a combination therof used in a study. For biological NAMs test system refers to the biological model utilised (e.g., cell line, primary cell, co-culture etc.). Test kits should also be considered as test systems (OECD GIVIMP, 2018).

Tiered (or hierarchical) testing approach: Describes an approach, where test methods are organised in a sequential (hierarchical) manner such that typically the outcome of one test would determine how to proceed. First tiers of such approaches use rather simple methods such as *in silico* predictions or (high-throughput) screening methods, while higher tiers involve more elaborated/ specific approaches. The highest tier could include *in vivo* tests.

Toxicodynamics: describes the dynamic interactions of a chemical (toxicant) with a biological target, explains how a chemical elicits a toxic response (in analogy in pharmacology the similar term pharmacodynamics explains the (inter-)actions of a drug).

Toxicokinetics: describes how a chemical is taken up, distributed, metabolised and finally excreted (in analogy pharmacokinetics describes these processes for drugs).

Transferability: The proof that a test method can also be reliably performed in independent, competent laboratories that before have not applied this method (also referred to as method-naïve laboratories).

Trueness: assessed by estimating the magnitude of the experimental bias and by determining if it is statistically significant. Bias can be determined by using certified reference materials (CRMs), by comparing the new method with a reference method or by making use of data coming from an interlaboratory comparison (common term for analytical methods).

Validated test method: A test method for which a multi-center validation study (overseen by a validation body) has been successfully completed to determine its relevance and reliability for a specific purpose.

Valid test method (or scientifically valid test method): A test method has proven to be sufficiently relevant and reliable for a specific purpose by other means than a multi-center validation study (for example by qualification).

Validation: The process by which the reliability and relevance of a particular approach, method, process or assessment is established for a defined purpose (OECD GD 34). Importantly, different validation approaches exist. The choice of the appropriate validation approach depends on the type of method, its intended use/ context of use and is furthermore also critically influenced by other factors such as the available resources. Test methods can be validated by different actors.

Working range: is defined by providing lower & upper concentrations, between which the analytical procedure provides an acceptable linearity, trueness and precision (common term for analytical methods).

# Abbreviations

| | |
|---|---|
| AC | Acceptance criteria |
| ADME | Absorption, Distribution, Metabolism and Excretion |
| AL | Apical |
| AOP | Adverse Outcome Pathway |
| BCOP | Bovine Corneal Opacity and Permeability test |
| BL | Basolateral |
| CEN | European Committee for Standardization |
| CLSM | Confocal Laser Scanning Microscopy |
| CME | Clathrin-mediated endocytosis |
| CMR | Carcinogenic, mutagenic, reprotoxic |
| DA | Defined Approach |
| DLS | Dynamic light scattering |
| DNT | Developmental Neurotoxicity |
| ED | Endocrine Disruption |
| EFSA | European Food Safety Authority |
| EM | Electron Microscopy |
| EMA | European Medicines Agency |
| EO | EpiOcular |
| EURL | European Union Reference Laboratories |
| EURL-ECVAM | European Union Reference Laboratory for Alternatives to Animal Testing |
| FDA | US Food and Drug Administration |
| GIT | Gastro-intestinal tract |
| GIVIMP | Guidance Document on Good In vitro Method Practices |
| GLP | Good Laboratory Practice |

| | |
|---|---|
| GUM | Guide to the expression of uncertainty in measurement (ISO/IEC Guide 98-1:2024) |
| HPLC | High Performance Liquid Chromatography |
| HTS | High-Throughput Screening |
| IATA | Integrated Approach to Testing and Assessment |
| ICH | International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use |
| ICP-MS | Inductively coupled plasma mass spectrometry |
| ILC | Interlaboratory Comparison |
| ISO | International Standardisation Organisation |
| KE | Key Event |
| LDH | Lactate dehydrogenase |
| LoD/ LoQ | Limit of Detection/ Limit of Quantification |
| LoI | Letter of Intent |
| LY | Lucifer Yellow |
| MAD | Mutual Acceptance of Data |
| M-cells | Microfold-cells |
| MDDT | Medical device development tools |
| MoA | Mode of Action |
| NAM | New Approach Methodology |
| NTA | Nanoparticle tracking analysis |
| NM | Nanomaterial |
| NP | Nanoparticle |
| OECD | Organisation for Economic Co-operation and Development |
| OECD GD | Guidance Document released by the OECD |
| OECD TG | Test Guideline released by the OECD |
| OECD WPMN | OECD Working Party on Manufactured Nanomaterials |

| | |
|---|---|
| Papp | Apparent permeability |
| PBK | Physiologically Based Kinetic |
| REACH | Regulation on the registration, evaluation, authorisation and restriction of chemicals |
| RDT | Repeated Dose Toxicity |
| RRL | Regulatory Readiness Level |
| SbD | Safe-by-Design |
| SOP | Standard Operating Procedure |
| spICP-MS | Single particle ICP-MS |
| SPSF | Standard project submission form |
| SSbD | Safe-and-Sustainable-by-Design |
| (Q)SAR | (Quantitative) Structure Activity Relationship |
| TEER | Transepithelial Electrical Resistance |
| TEM | Transmission Electronic Microscopy |
| TG | Test Guideline |

# Appendix A: Criteria for *in chemico* and analytical NAMs

**Analytical and *in chemico* NAMs -** Letter of Intent

Please note that the general information is identical to Section 3.2.2.

| Sub-Section | Criterion | Guidance |
|---|---|---|
| Assay | What is actually measured? | Specific properties measured (e.g. which specific property of the NM that is addressed). Indicate the specific conditions under which the measurement is performed. Also describe here any potential interference, confounder, reference/normalization procedure that may be used. |
| | How is it measured? | Analytical technique used (e.g. EM, DLS, mass spectrometry) |
| Relevance (for context-of-use) | Relevance of the measurement? | What is the relevance of the measurand in the context of risk assessment |
| | How does it fit in the regulatory framework? | Is it a screening assay? Is it part of a test battery or chain of models? Which information need is addressed? |

**Analytical and *in chemico* NAMs -** Qualification Plan

| Sub-Section | Criterion | Guidance |
|---|---|---|
| **NAM Description, consider to I) Set-up (all elements), II) application phase, III) evaluation phase** | | |
| Detailed SOP(s) | Covering all phases/ steps (set-up of the method, application, evaluation phase) | Firstly, all elements of the method (set-up) have to be described. Secondly, all steps of the application must be written down in detail with highest possible clarity to fully understand/ evaluate a method (including e.g. test item preparation). Finally, the data evaluation needs to be described. This generally requires several SOPs for the different phases/steps.<br><br>Custom-made instrument(s) or specialised equipment(s), which are not standard in any laboratory |
| **Reliability** | | |
| Reliability | Linearity | As demonstrated by calibration |
| | Working Range | Lower & Upper limits, between which the analytical procedure provides an acceptable linearity, trueness and precision |
| | LoD/ LoQ | Lowest concentration or size that can be reliably detected (LoD)/ reliably quantified (LoQ) |

| Sub-Section | Criterion | Guidance |
|---|---|---|
| | Trueness | Use of certified reference materials, comparison to standard methods |
| | Selectivity | Ability to correctly measure without interferences from other analytes |
| | Repeatability | Ability of the method to generate the same results over a short time interval under identical conditions. It is recommended to measure at least ten subsamples in one run or three subsamples over five different days |

**Analytical and *in chemico* NAMs -** Qualification Proposal

| Sub-Section | Criterion | Guidance |
|---|---|---|
| **NAM Description, consider to I) Set-up (all elements), II) Application phase, III) Evaluation phase** | | |
| Overarching | Acceptance criteria | XXX |
| | Known limitations | XXX |
| | Trouble shooting | XXX |
| **Reliability and Relevance (for context-of-use)** | | |
| Reliability | (Inter)mediate Precision | Agreement between the results from within-laboratory variations due to random events that might occur when using the method such as testing on different days, involving different analysts, using different equipment and/or under different calibration. At least the variation of one of these factors has to be investigated. These studies are carried out according to ISO 5725-3 (2023). |
| | Transferability | Minimum one other lab should have applied the method. How comparable is outcome? Which amount of training was required? |
| Relevance | Data from orthogonal methods? | XXX |

# Appendix B: Readiness criteria for *in silico* QSAR NAMs

***In-silico* QSAR NAMs -** Letter of Intent

Please note that the general information is identical to Section 3.2.2.

| Sub-Section | Criterion | Guidance |
|---|---|---|
| Model description | Data Sources | Sources of the experimental data that have been used to set-up the model, including all details on the associated data set(s) |
| Algorithm | Unambiguous algorithm | Description of the algorithm/ software, Description of all inputs and settings (of the software, if applicable) |
| Relevance (for context-of-use) | Relevance of the model | What is the relevance of the model in the context of risk assessment? |
| | How does it fit in the regulatory framework? | Is it part of a test battery or chain of models? Which information need is addressed? |

***In-silico* QSAR NAMs -** Qualification Plan

| Sub-Section | Criterion | Guidance |
|---|---|---|
| **NAM Description, consider to I) Set-up (all elements), II) Application phase, III) Evaluation phase** | | |
| Detailed Descriptions of the QSAR model | XXX | Detailed description of the QSAR model (how was it set up, based on which data set, which algorithm, how was it trained etc).<br>Detailed description of how the QSAR model shall be applied.<br>Elaboration on how the outcome can be used.<br><br>Follow the details described in the QSAR assessment framework (OECD, 2023c) |
| **Reliability** | | |
| Reliability | Verification | Are all input/ output parameters and their units correct? |
| | Quality of input parameters | Please elaborate on the quality of the data that has been used to set-up the model |
| | Sensitivity analysis | Identified the model parameters that have a significant influence on model outputs |
| | Reproducibility | Is it verified that the outcome is reproducible |
| | Standard error | XXX |
| | Uncertainty analysis | What is the associated uncertainty |

***In-silico* QSAR NAMs -** Qualification Proposal

| Sub-Section | Criterion | Guidance |
|---|---|---|
| **NAM Description, consider to I) Set-up (all elements), II) Application phase, III) Evaluation phase** | | |
| Variability | XXX | XXX |
| Acceptance criteria | XXX | XXX |
| Known limitations | XXX | XXX |
| Trouble shooting | XXX | XXX |
| **Relevance** | | |
| Relevance | Applicability Domain/ Predictive Capacity | Clear indication which chemicals/ type of NMs the model is applicable |
| | Goodness-of-fit | How large was training data set? |
| | Biological Plausibility | Please explain why/ how the model is relevant (e.g. mechanistic interpretation for QSAR model, knowledge on uptake/ distribution pathways) |

# Appendix C: Readiness criteria for *in silico* PBK NAMs

***In-silico* PBK NAMs -** Letter of Intent

Please note that the general information is identical to Section 3.2.2.

| Sub-Section | Criterion | Guidance |
|---|---|---|
| Model description | Description of the model | How many compartments? Perfusion- or permeability limited? |
| Model parameters | Modelling Parameters | Which parameters are needed to run the model? |
| | Software | Commercial PBK platform/ software used? |
| | Implementation | Does the model code express the mathematical relationship? |
| Relevance (for context-of-use) | Relevance of the model | What is the relevance of the model in the context of risk assessment? |
| | How does it fit in the regulatory framework? | Which information need is addressed? How shall it be applied in risk assessment? |

***In-silico* PBK NAMs -** Qualification Plan

| Sub-Section | Criterion | Guidance |
|---|---|---|
| **NAM Description, consider to I) Set-up (all elements), II) Application phase, III) Evaluation phase** | | |
| Detailed Descriptions of the PBK model | XXX | Detailed description of the QSAR model (how was it set up, based on which data set, which algorithm, how was it trained etc.)<br>Detailed description of how shall the QSAR model be applied.<br>Elaboration on how the outcome can be used.<br><br>Follow the details described in the PBK guidance (OECD, 2021b). |
| **Reliability** | | |
| Reliability | Verification | Is the chemical mass balance respected at all times? Is the cardiac output equal to the sum of blood flow rates to the tissue compartments? |
| | Quality of input parameters | Please elaborate on the quality of the data that has been used to set-up the model |
| | Sensitivity analysis | Identified the model parameters that have a significant influence on model outputs. |
| | Reproducibility | Is it verified that the outcome is reproducible |
| | Standard error | |
| | Uncertainty analysis | What is the associated uncertainty |

***In-silico* PBK NAMs -** Qualification Proposal

| Sub-Section | Criterion | Guidance |
|---|---|---|
| **NAM Description, consider to I) Set-up (all elements), II) Application phase, III) Evaluation phase** | | |
| Variability | XXX | XXX |
| Acceptance criteria | XXX | XXX |
| Known limitations | XXX | XXX |
| Trouble shooting | XXX | XXX |
| **Relevance** | | |
| Relevance | Applicability Domain/ Predictive Capacity | Clear indication which chemicals/ type of NMs the model is applicable |
| | Goodness-of-fit | Are there any empirical data (e.g. in vivo kinetic data) to demonstrate the goodness-of-fit |
| | Biological Plausibility | Please explain why/ how the model is relevant (e.g. knowledge on uptake/ distribution pathways) |

# Annex A – Readiness Criteria for NAM Qualification

This Annex is an excel table containing the detailed criteria for evaluating method readiness for biological NAMs and is published in the supporting documentation section of this report.