EVIDENCE-BASED TOXICOLOGY
2023, VOL. 1, NO. 1, 2232415
https://doi.org/10.1080/2833373X.2023.2232415

Taylor & Francis
Taylor & Francis Group

RESEARCH ARTICLE

OPEN ACCESS Check for updates

# Protocol for designing INVITES-IN, a tool for assessing the internal validity of *in vitro* studies

Camilla Svendsen[a,b], Paul Whaley[a,c], Gunn E. Vist[a,d], Trine Husøy[a,e], Anna Beronius[f], Emma Di Consiglio[g], Ingrid Druwe[h], Thomas Hartung[i,j], Vasiliki I. Hatzi[k], Sebastian Hoffmann[l,m], Carlijn R. Hooijmans[n], Kyriaki Machera[k], Joshua F. Robinson[o], Erwin Roggen[p], Andrew A. Rooney[q], Nicolas Roth[r,s], Eliana Spilioti[k], Anastasia Spyropoulou[k], Olga Tcheremenskaia[g], Emanuela Testai[g], Mathieu Vinken[t] and Gro H. Mathisen[a]

[a]Norwegian Scientific Committee for Food and Environment, Norwegian Institute of Public Health, Oslo, Norway; [b]Department of Chemical Toxicology, Norwegian Institute of Public Health, Oslo, Norway; [c]Lancaster Environment Centre, Lancaster University, Lancaster, UK; [d]Division for Health Services, Norwegian Institute of Public Health, Oslo, Norway; [e]Department of Food Safety, Norwegian Institute of Public Health, Oslo, Norway; [f]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden; [g]Environment & Health Department, Italian National Institute of Health (ISS), Rome, Italy; [h]United States Environmental Protection Agency, Office of Research and Development, Center for Public Health and Environmental Assessments, Research Triangle Park, NC, USA; [i]Center for Alternatives to Animal Testing (CAAT), Johns Hopkins University, Bloomberg School of Public Health, Baltimore, MD, USA; [j]CAAT Europe, University of Konstanz, Konstanz, Germany; [k]Laboratory of Toxicological Control of Pesticides, Scientific Directorate of Pesticides' Control and Phytopharmacy, Benaki Phytopathological Institute, Kifissia, Greece; [l]Evidence-Based Toxicology Collaboration (EBTC), Johns Hopkins University, Bloomberg School of Public Health, Baltimore, MD, USA; [m]SEH consulting + services, Paderborn, Germany; [n]Department of Anesthesiology, Pain and Palliative Care, Radboud University Medical Centre, Nijmegen, Netherlands; [o]Department of Obstetrics, Gynecology & Reproductive Sciences, University of California, San Francisco (UCSF), CA, USA; [p]3Rs Management and Consulting ApS, Lyngby, Denmark; [q]Division of Translational Toxicology, National Institute of Environmental Health Sciences, Research Triangle Park, NC, USA; [r]Department of Pharmaceutical Sciences, University of Basel, Basel, Switzerland; [s]Swiss Centre for Applied Human Toxicology (SCAHT), Basel, Switzerland; [t]Department of Pharmaceutical and Pharmacological Sciences, Vrije Universiteit Brussel, Brussel, Belgium

**ABSTRACT**

This protocol describes the design and development of a tool for evaluation of the internal validity of *in vitro* studies, which is needed to include the data as evidence in systematic reviews and chemical risk assessments. The tool will be designed specifically to be applied to cell culture studies, including, but not restricted to, studies meeting the new approach methodology (NAM) definition. The tool is called INVITES-IN (IN VITro Experimental Studies INternal validity).

In this protocol, three of the four studies that will be performed to create the release version of INVITES-IN are described. In the first study, evaluation of existing assessment tools will be combined with focus group discussions to identify how characteristics of the design or conduct of an *in vitro* study can affect its internal validity. Bias domains and items considered to be of relevance for *in vitro* studies will be identified. In the second study, group agreement on internal validity domains and items of importance for *in vitro* studies will be identified via a modified Delphi methodology. In the third study, the draft version of the tool will be created, based on the data on relevance and importance of bias domains and items collected in Studies 1 and 2. A separate protocol will be prepared for the fourth study, which includes the user testing and validation of the tool, and collection of users' experience.

**Abbreviations:** NAM: new approach methodologies; PG: project group; SAG: scientific advisory group

**ARTICLE HISTORY**
Received 9 March 2023
Revised 19 June 2023
Accepted 19 June 2023

**KEYWORDS**
Cell culture; NAMs; next generation risk assessment; risk of bias

## 1. Introduction

### 1.1. Evaluation of internal validity

This protocol describes the design and development of a tool for evaluation of the internal validity of *in vitro* studies. Internal validity is the extent to which a study (methodological design, methods and data analysis) is free from bias, where bias is 'systematic error, or deviation from the truth, in results' (Cochrane Collaboration 2005). A test performed *in vitro* ('in the glass') means

**CONTACT** Gro H. Mathisen gro.haarklou.mathisen@vkm.no

This manuscript was accepted for publication by the handling editor Olwenn Martin after 1 round of editorial evaluation and 1 round of peer-review evaluation. The evaluation reports for the manuscript can be found at DOI 10.5281/zenodo.7726071. Preprint versions of the manuscript and author responses to comments in the evaluation reports can be found at DOI 10.5281/zenodo.7633626

that it is done outside of a living organism and it usually involves isolated tissues, organs or cells (ECHA 2023). The tool is called INVITES-IN (IN VITro Experimental Studies INternal validity).

Methods to generate evidence for regulatory toxicology are shifting from classical animal experiments to new approach methodologies (NAMs). The European Chemicals Agency and the U.S. Environmental Protection Agency define NAMs as any technology, methodology, approach or combination that can provide information on chemical hazard and risk assessment without the use of animals, including *in silico*, *in chemico*, *in vitro* and *ex vivo* approaches (ECHA 2016; EPA 2018). According to the European Food Safety Authority (EFSA), the term NAMs is used to make reference to any non-animal-based approach that can be used to provide toxicological information in the context of hazard/risk assessments (EFSA et al. 2022).

As part of the gradual incorporation and transition toward the use of NAMs, including *in vitro* studies, a framework for evidence-based use of NAMs in toxicological research and chemical risk assessment is required. Such a framework should ultimately incorporate at least the following principles:

1. Result in identification of all relevant NAM-generated evidence relating to the research question addressed in a systematic review or risk assessment.
2. Provide for the evaluation of the internal validity of NAM studies (propensity for systematic error due to how the study is designed and conducted).
3. Provide for the evaluation of the external validity of NAM studies (the degree to which results of a study can be translated/generalised to human adverse health effects).
4. Contribute to objectivity, robustness, transparency and reproducibility in the hazard identification and characterisation process.
5. In its approach to normalising and structuring the description and analysis of NAMs, contribute to progress in the extent to which research data conform to FAIR (Findable, Accessible, Interoperable and Re-usable) principles of open science.

Systematic review and evidence-based toxicology principles should be implemented in all parts of the framework, and it should be generic and usable across different regulatory sectors such as food safety, cosmetic ingredient safety, etc. Principles for incorporating evidence from NAMs into risk assessments and a framework for the evaluation of skin sensitisation have been developed for cosmetic ingredients (Dent et al. 2018; Gilmour et al. 2020). Methods for incorporation of mechanistic studies as supporting evidence in hazard and/or risk assessment is included in the U.S. NTP OHAT handbook for systematic reviews, the ORD staff handbook for developing IRIS assessments, and the draft TSCA interpretation of systematic review methods to support chemical risk evaluations (EPA 2022, 2023; NTP OHAT 2019). However, there is currently no complete framework for evidence-based chemical risk assessment that integrates NAMs to facilitate the transition from use of animals to the use of NAMs in chemical risk assessments.

'Next generation risk assessment in practice' is a project in the European Partnership for the Assessment of Risks from Chemicals (PARC). PARC aims to develop next generation chemical risk assessment to advance research, share knowledge and improve skills, protecting human health and the environment. The present project is included in the task focusing on facilitating regulatory acceptance and use of NAMs. PARC is a 7-year partnership under Horizon Europe, including close to 200 institutions from 28 countries working in the areas of the environment or public health, and 3 EU authorities (PARC 2023). With the 'Next generation risk assessment in practice' project, we aim to contribute to the development of a framework for evidence-based use of data generated by *in vitro* studies in human health hazard identification and characterisation by creating tools and guidances. A webpage giving an overview of the planned work in the 'Next generation risk assessment in practice' project has been created (VKM 2023). The first step in our PARC project is to develop a tool for evaluation of internal validity for *in vitro* studies. The next steps, all focusing on *in vitro* studies, will be the development of a tool for evaluation of external validity, creation of guidance for evaluation of certainty in the evidence, and creation of guidance for the identification of point of departure and the uncertainty in the point of departure. We chose to start focusing on creation of tools for validity assessment, as validity assessment is one of the critical steps in the systematic review process. Further, we chose to start focusing on *in vitro* models as there is a general agreement that these are important as replacement for animal studies to provide information for hazard/risk assessment (ECHA 2016; EFSA et al. 2022; EPA 2018) in a wider integrating approach. It has been suggested that *in vitro* models could be more suitable than animal models for the prediction of toxicity. For example, *in vitro* data did predict liver toxicity caused by the drug troglitazone whereas neither

published animal nor human studies were able to accurately predict the hazard (Dirven et al. 2021).

Several *in vitro* study designs exist; however, we have chosen only to focus on cell culture studies (meaning studies using cells derived from multicellular organisms). This delimitation is mainly due to feasibility, especially concerning the user testing, where the number of user testing participants will have to be very large to be able to test that the tool works on all types of *in vitro* study designs.

The implementation of this tool might be of help to improve the inclusion of NAMs in the chemical risk assessment process and facilitate regulatory uptake, with a focus on risk assessors' daily practice and workflow.

While many tools have been created for assessing *in vitro* studies, there is a priori lack of consensus on developing a tool with the application of rigorous methods. We therefore aim to address this situation by using methods that ensure we are building on prior work, with a degree of rigor consistent with our intent to provide an authoritative assessment tool. We also intend to use the findings of INVITES-IN to prepare guidance on the design and conduct of *in vitro* studies that will help researchers minimise and/or transparently identify potential biases in their studies.

### 1.2. Objective

The aim of this project is to create INVITES-IN, a tool for evaluating the internal validity of *in vitro* studies. The INVITES-IN tool will be designed specifically to be applied to cell culture models (e.g., cell lines, primary cell models, co-cultures, monolayer and 3-D cell models systems) treated with a single-chemical substance exposure, measuring any outcome. We anticipate that the tool will be applicable (potentially with modification) to other *in vitro* study designs or other NAMs such as organ-on-a-chip, *in ovo, fish embryos, ex vivo*, *in chemico,* etc., and chemical mixture studies, but this will not be addressed in this study.

To contribute to its usability, INVITES-IN will be accompanied by instructions to guide the user through the evaluation of internal validity of *in vitro* studies step-by-step. While there is good empirical evidence from several domains that certain features of how a study is designed, conducted and analysed can introduce bias, it is usually not possible to determine how much bias a given feature has introduced on any specific occasion (Savović et al. 2012). INVITES-IN therefore follows conventional guidance (Boutron et al. 2022; Frampton et al. 2022) in being designed to differentiate studies with relatively higher risk of bias from studies with relatively lower risk of bias.

### 1.3. Project governance

The development of INVITES-IN is part of the PARC project 'Next generation risk assessment in practice' [Project 101057014 – PARC]. A project group (PG) has been established with the responsibility for developing and implementing the tool for evaluation of internal validity of *in vitro* studies. The project is led by the Norwegian Institute of Public Health represented by the Norwegian Scientific Committee for Food and Environment (Norway). The project partners are Benaki Phytopathological Institute (Greece), Istituto Superiore di Sanità (Italy) and the University of Basel (Switzerland).

A scientific advisory group (SAG) consisting of experts in systematic review principles, chemical risk assessment, toxicology, NAMs and/or methods for tool development, several of whom have been directly involved in developing approaches to assessing the validity of *in vitro* studies, has been established. The SAG gives strategic guidance and support to the PG and share information about ongoing projects addressing similar questions to ensure that the outcome of this project complements and builds on the work of others and thereby creates synergies and avoids duplication of efforts.

## 2. Materials and methods

### 2.1. Study design

#### 2.1.1. An overview of the creation of INVITES-IN

The method for creating INVITES-IN will follow the general framework for developing quality assessment tools suggested by Whiting et al. (2017). This is a broad framework of general principles rather than a tightly prescribed standard but gives the general structure of our approach. Four studies will be performed to create INVITES-IN (Figure 1). This protocol describes Studies 1, 2 and 3, and the timeline is shown in Figure 2. A separate protocol will be prepared for Study 4.

The tool will consist of signalling questions and criteria for reaching risk-of-bias judgments for each signalling question. Criteria are the issues that have to be fulfilled to avoid bias. Signalling questions are questions that the users of the tool answer in order to determine whether the criteria have been fulfilled. The technical solution for the tool has not yet been decided; however, we intend to make an online tool.

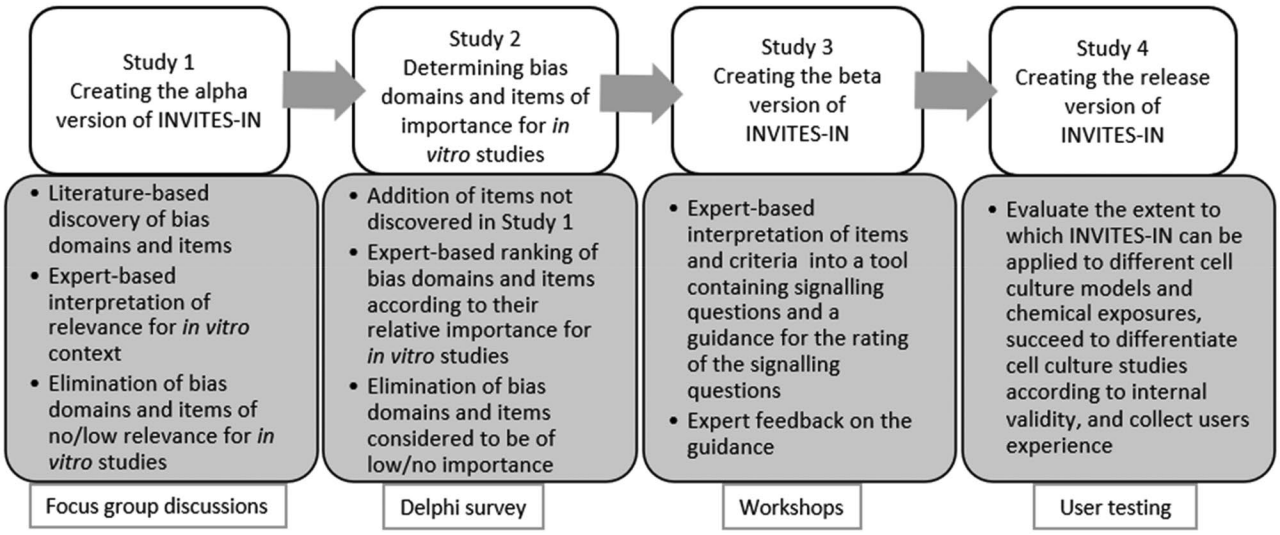


**Figure 1.** An overview of the four studies that will be performed to create the release version of INVITES-IN.


**Figure 2.** An overview of the 2023–2024 timeline for the creation of the beta version of INVITES-IN.

The target group for the use of the tool (i.e., end-users) includes *in vitro* scientists and risk assessors conducting literature reviews in hazard assessments/safety evaluations, which could be part of a chemical risk/safety assessment, a systematic review or both, for regulatory or research purposes.

To get the input we need to develop the tool, we aim to recruit participants experienced with *in vitro* research that are representative for the end-users. For the Studies 1 and 3, we aim to recruit some participants also having experience with systematic reviews, some also having experience with chemical risk assessment, and some having no experience with systematic reviews or chemical risk assessments. For Study 2, we consider it critical that all participants have systematic review experience, as this is the study where the importance of different internal validity items will be ranked. Previous experience with evaluation of internal validity is considered important to be able to rank importance of different internal validity items. All groups of end-expected users are covered by the networks of the PG and the SAG. Potential participants will therefore be identified through nomination by PG and SAG members, who will be requested to nominate three potential participants. For each nominated participant, an overview of their scientific expertise and experience, affiliation, geographical location and gender will be prepared. From the pool of nominated participants, PG will select participants that will be invited. In the selection process, PG will ensure diversity among the participants by including scientists from different fields having different professional backgrounds and experience with different cell culture models, covering a variety of geographical locations, and having an even gender distribution. In each focus group, all participants should be affiliated with different institutions, located in at least four different countries. This way we will avoid having an overrepresentation of focus group participants from a few institutions or from a too limited number of countries. We consider that this described process will make it possible to carry out the recruitment without it being an overly

time-consuming process, and at the same time secure sufficient diversity in the group of participants.

The tasks and workload for the participants, the outcome of their contribution and the participant eligibility criteria, are shown in Figure 3 and Table 1. Note that it is not expected that the same persons participate in all studies. It is planned that the persons participating in Study 1 will be also invited to participate in Study 3.

For all three studies, the potential participants will receive information about the project when they are contacted by email, and participants that accept the invitation will be requested to complete a declaration of interest form. The PG will evaluate the declaration of interest forms, focusing mainly on identification of potential conflicts of interest that may interfere with the participants' contribution and role in the focus group discussion.

Previous studies report average or median time for the assessment of RoB of a study to range from 20 to 40 min (Eick et al. 2020; Momen et al. 2022). We intend to keep the time needed for assessment of one cell culture study within this range.

All data analyses will be done by the PG members. All raw data from each study will be anonymised and made available as supplementary to the respective publications.

### *2.1.2. Ethical review*

Ethical approval has been given by the Norwegian Institute of Public Health.

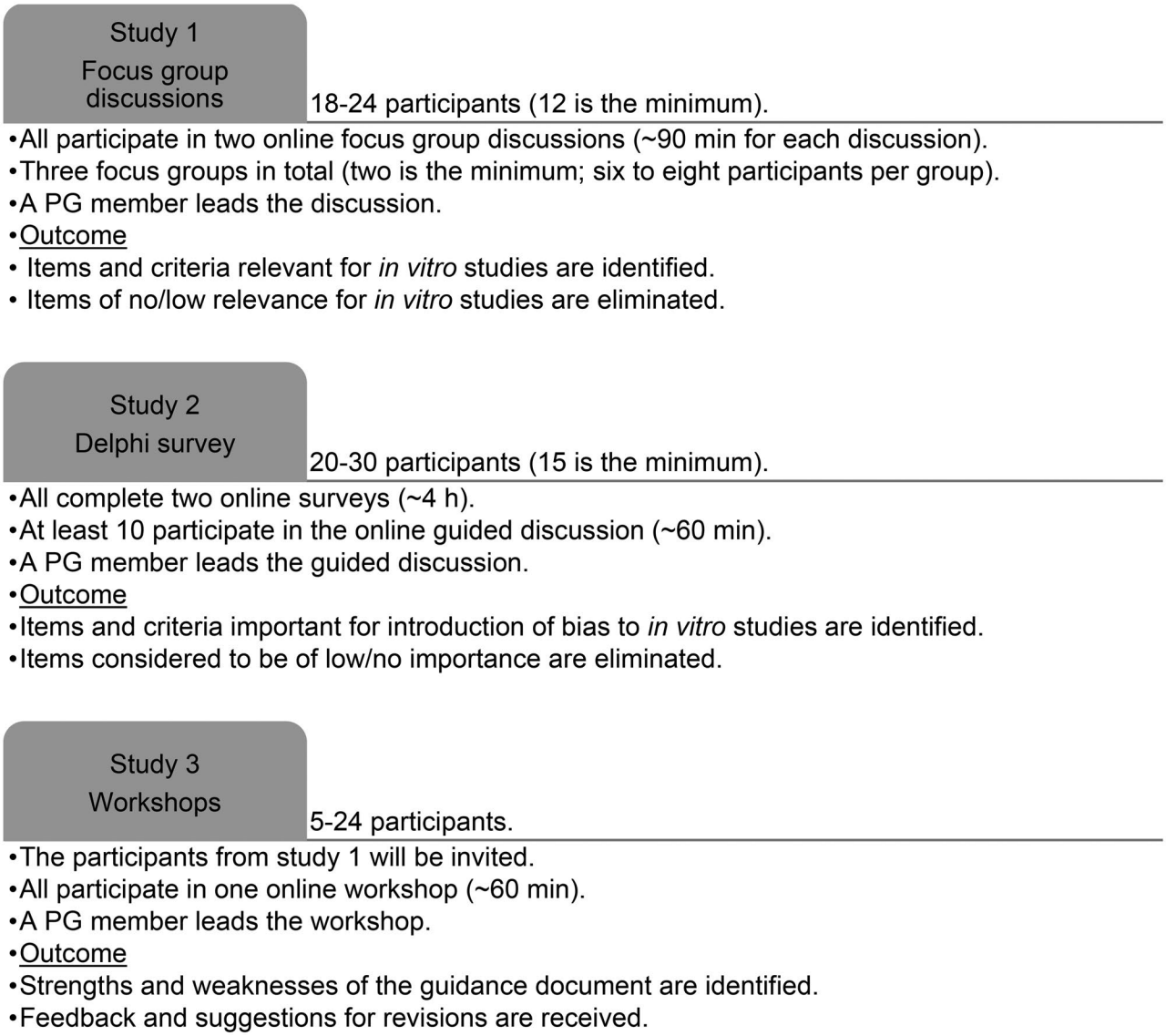


**Figure 3.** Participants' tasks and workload in Studies 1–3, and the outcome of their contribution.

**Table 1.** An overview of the criteria for participation in Studies 1–3.

| Selection of participants | | Study 1 | Study 2 | Study 3 |
|---|---|---|---|---|
| Scientific experience and expertise | *In vitro* models | x | | x |
| | *In vitro* models AND chemical risk assessment | x | | x |
| | *In vitro* models AND systematic review methods | x | x | x |
| | *In vitro* models AND experienced with the development of relevant guidance documents for chemical risk assessors | x | | x |
| Balancing factors | Academia | x | x | x |
| | Governmental institutions (including risk assessment institutions and research institutes) | x | x | x |
| | Private sector research institutions | x | x | x |
| | Gender distribution | x | x | x |
| | Demographic distribution | x | x | x |
| | Regional distribution | x | x | x |
| Academic level | Post-doctoral level or higher | x | x | x |
| Language | English, level B1 or higher | x | x | x |

**Table 2.** An overview of Study 1.

| Phase | Task | Responsible |
|---|---|---|
| Plan | Prepare the list of bias domains and items. | Project group |
| | Create questions for the focus group discussions. | |
| | Define inclusion criteria for focus group participants. | Project group and scientific advisory group |
| | Nominate and recruit focus group participants fulfilling the inclusion criteria. | |
| Actions | Carry out the focus group discussions. | Project group |
| | Analyse results and prepare the final report. | |
| Result | Bias domains and items of relevance for *in vitro* studies are identified and included in the alpha version of the tool. | Project group |

## 2.2. Study 1: Creating the alpha version of the tool

### 2.2.1. Introduction and objective

The objective of Study 1 is to create a straw-man or alpha version of INVITES-IN that can be further developed via a modified Delphi process (see Section 2.3.2 for description). In Study 1, a list of characteristics of the design, conduct and analysis of an *in vitro* study that can introduce bias into its results or findings will be compiled, organised thematically and then interpreted into a draft set of structured signalling questions that constitute the alpha version of INVITES-IN.

The knowledge goal is to have the expert interpretations of the relevance of bias domains and items for *in vitro* studies.

A pilot focus group discussion was arranged to get an impression of the time needed for the focus group discussions, to test the technical functions and to get feedback on factors related to the presentation of questions and the use of examples that may be of importance to conduct successful focus group discussions.

### 2.2.2. Method

We will include three focus groups with six to eight participants in each group (Figure 3).

An overview of the workflow and the responsibilities in Study 1 are given in Table 2.

***2.2.2.1. Identifying relevant bias domains and items.*** A list of bias domains and items of potential relevance for *in vitro* studies will be prepared using several literature sources. This list will serve as a starting point for the creation of INVITES-IN and provide the basis for the focus group discussions. The literature sources are as follows: two systematic reviews on validity tools for *in vitro* models (Tran et al. 2021; Whaley, Hooijmans, and Wattam in preparation), a publication on study sensitivity that includes assessment items that may relate to internal validity but may not be included in other tools (Cooper et al. 2016) and tools for evaluation of risk of bias (EPA 2022; NTP OHAT 2015, 2019; Roth, Zilliacus, and Beronius 2021; Sterne et al. 2019).

***2.2.2.2. Focus group participants.*** Eligible focus group participants will be scientists with or without systematic review experience that are active in the field of *in vitro* research in academia, governmental institutions (including risk assessment institutions and research institutes) or private research institutes, at post-doctoral level or higher, and level B1 English speakers (Table 1). PG and SAG will nominate participants. We aim to have an equal gender distribution, a reasonable demographic and regional distribution, and a group size of six to eight participants as this group size is recommended to generate diverse ideas but not so many participants that they do not have a chance to share perspectives (Krueger et al. 2001). The minimum number of participants in a focus group is considered

**Table 3.** Bias domains with approved definitions in the SEVCO (FEvIR Platform Version 0.80.0, 06.12.2022).

| Bias domain | Definition | SEVCO code reference |
|---|---|---|
| Selection bias | A bias resulting from methods used to select subjects or data, factors that influence initial study participation, or differences between the study sample and the population of interest | SEVCO:00002 |
| Confounding covariate bias | A situation in which the effect or association between an exposure and an outcome is distorted by another variable. For confounding covariate bias to occur, the distorting variable must be (1) associated with the exposure and the outcome, (2) not in the causal pathway between exposure and outcome and (3) unequally distributed between the groups being compared. | SEVCO:00016 |
| Performance bias | A bias resulting from differences between the received exposure and the intended exposure. | SEVCO:00017 |
| Attrition bias | A bias due to the absence of expected participation or data collection after selection for study inclusion. | SEVCO:00019 |
| Detection bias | A bias due to distortions in any process involved in the determination of the recorded values for a variable. | SEVCO:00020 |
| Analysis bias | A bias related to the analytic process applied to the data. | SEVCO:00021 |
| Reporting bias | A bias due to distortions in the selection or representation of information in study results or research findings. | SEVCO:00023 |
| Early study termination bias | A bias due to the decision to end the study earlier than planned. | SEVCO:00370 |

to be four. All participants in a focus group will be affiliated with different institutions in an attempt to achieve variation in input and perspective, and they should be working with a variety of *in vitro* models to cover a wide range of experimental systems. No compensation is offered for the participation, and participants will not be offered co-authorship.

Potential focus group participants will be contacted via email. They will receive a document with information about the project, the purpose of the focus groups and the focus group discussions, that the use of information learned in the meeting will not allow for identification of the focus group participants, the withdrawal procedure, the financial source and the approximate time for the focus group meeting. Focus group participants must actively confirm their consent by email.

We aim to have three different focus groups (Krueger et al. 2001); however, two groups are considered to be the minimum. All groups will be presented with the same information and questions, although the direction in which discussion is steered may depend on how comprehensively previous focus groups were able to cover each issue. The need for including an additional group will be discussed if new insights are presented during the meetings, or if areas needing discussion were not addressed.

***2.2.2.3. Focus group discussion.*** We plan to have two group discussions per focus group. The second meeting will be cancelled if considered not to be needed. The discussions will be carried out as online meetings and will be recorded. A PG member will act as a focus group moderator and lead the discussions in the meeting, and another PG member will handle the logistics (the assistant moderator).

The complete list of identified bias domains and items will be the starting point for the focus group discussions. The discussions will be facilitated with a view to addressing two questions (numbering is for referencing purposes and the questions will not necessarily be presented in this order):

1. Are there any gaps in the identified domains or items that could influence systematic error in an *in vitro* study?
2. What characteristics of the design, conduct or analysis of an *in vitro* study could introduce systematic error into its results or findings?

Question (1) will be addressed both by asking directly and inferred from analysis of the discussion (see Section 2.2.2.4). Question (2) will be directly asked.

Discussion relating to questions (1) and (2) will be structured in terms of the bias domains defined in the Scientific Evidence Code System (SEVCO) (Table 3) (Alper et al. 2021b). The SEVCO domains are chosen because they are consistent with the bias domains of Whaley, Hooijmans, and Wattam (in preparation) and the OHAT tool (NTP OHAT 2019) but represent a more recent normalised list of bias categories derived from a robust grounding and consensus process (Alper et al. 2021a). These definitions are developed for human studies, and the relevance for *in vitro* studies will be discussed in the focus groups. We acknowledge that not all bias domains presented in Table 3 may be of relevance for *in vitro* studies. However, we will include all bias domains with approved SEVCO definitions in the focus group discussions in order to collect expert feedback on the relevance for *in vitro* studies. SEVCO draft bias domains that have not been approved are not listed. Participants may suggest additional bias domains.