Assessment of (Q)SAR predictions and results

The New OECD (Q)SAR Assessment Framework: Details and Examples

ASCCT-ESTIV Webinar, 30 October 2023

Andrea Gissi, PhD
Scientific Officer
European Chemicals Agency

The views expressed in this presentation are those of the author and do not necessarily reflect the official position of the European Chemicals Agency

Table of content

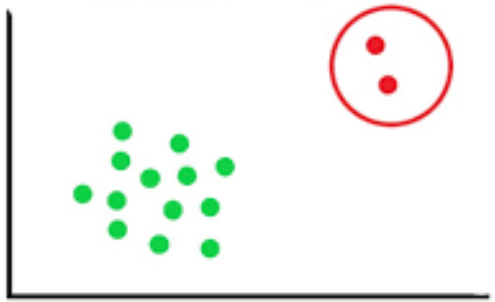
(Q)SAR Assessment Framework

Table of contents

Assessment of individual predictions

Valid (Q)SAR model ≠ Valid (Q)SAR result

- → The use of (Q)SARs is allowed in many chemical regulations
- → OECD (Q)SAR principles from 2004 cover the scientific validity of **(Q)SAR models**
- → The use of a valid (Q)SAR model does not guarantee the validity of each of its results
- → Need to establish **principles to assess individual results** and a systematic and harmonised **assessment framework** for (Q)SAR models and predictions

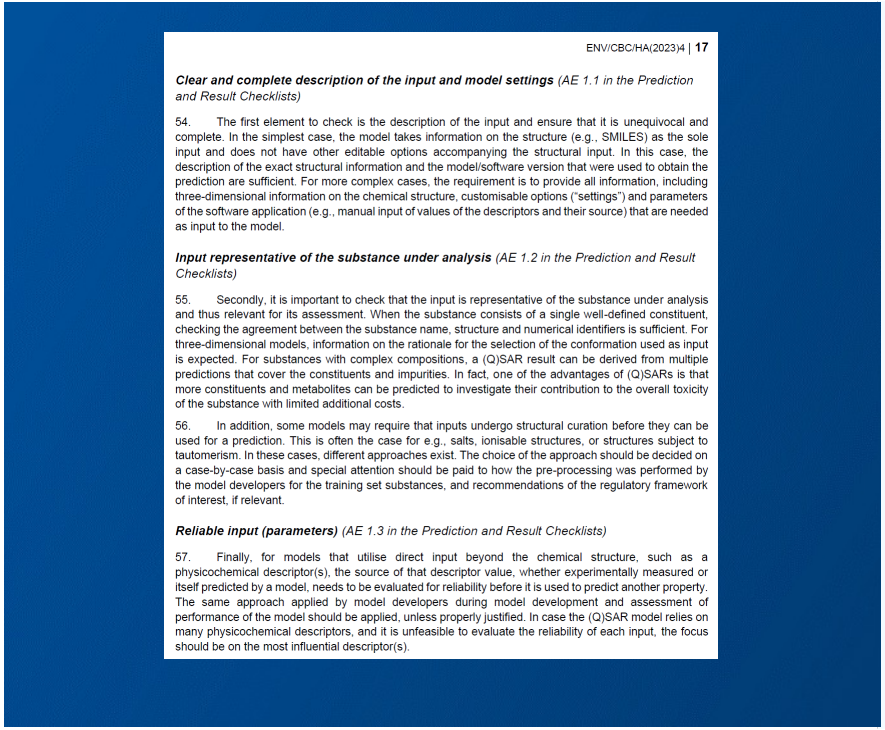


Principles for the assessment of (Q)SAR predictions

- Four new OECD principles for evaluating (Q)SAR predictions and results based on multiple predictions:
  1. **Correct input**
  2. **Substance within applicability domain**
  3. **Reliable prediction**
  4. **Outcome fit for purpose**

- For a result based on multiple predictions, each prediction is assessed individually, and then an additional evaluation step is dedicated to the final result

Guidance for the assessment of (Q)SAR predictions



Clear and complete description of the input and model settings *(AE 1.1 in the Prediction and Result Checklists)*

54. The first element to check is the description of the input and ensure that it is unequivocal and complete. In the simplest case, the model takes information on the structure (e.g., SMILES) as the sole input and does not have other editable options accompanying the structural input. In this case, the description of the exact structural information and the model/software version that were used to obtain the prediction are sufficient. For more complex cases, the requirement is to provide all information, including three-dimensional information on the chemical structure, customisable options ("settings") and parameters of the software application (e.g., manual input of values of the descriptors and their source) that are needed as input to the model.

Input representative of the substance under analysis *(AE 1.2 in the Prediction and Result Checklists)*

55. Secondly, it is important to check that the input is representative of the substance under analysis and thus relevant for its assessment. When the substance consists of a single well-defined constituent, checking the agreement between the substance name, structure and numerical identifiers is sufficient. For three-dimensional models, information on the rationale for the selection of the conformation used as input is expected. For substances with complex compositions, a (Q)SAR result can be derived from multiple predictions that cover the constituents and impurities. In fact, one of the advantages of (Q)SARs is that more constituents and metabolites can be predicted to investigate their contribution to the overall toxicity of the substance with limited additional costs.

56. In addition, some models may require that inputs undergo structural curation before they can be used for a prediction. This is often the case for e.g., salts, ionisable structures, or structures subject to tautomerism. In these cases, different approaches exist. The choice of the approach should be decided on a case-by-case basis and special attention should be paid to how the pre-processing was performed by the model developers for the training set substances, and recommendations of the regulatory framework of interest, if relevant.

Reliable input (parameters) *(AE 1.3 in the Prediction and Result Checklists)*

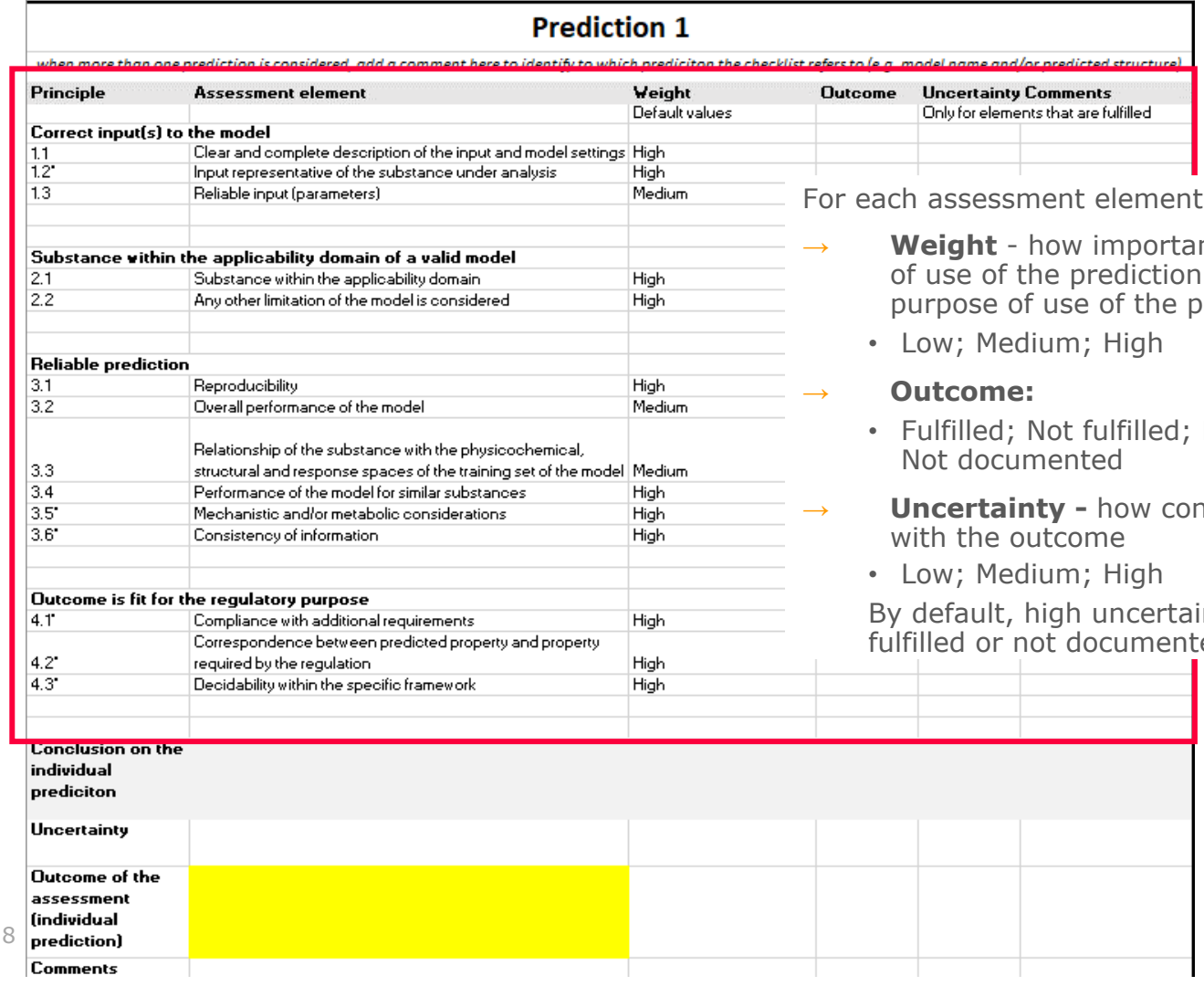
57. Finally, for models that utilise direct input beyond the chemical structure, such as a physicochemical descriptor(s), the source of that descriptor value, whether experimentally measured or itself predicted by a model, needs to be evaluated for reliability before it is used to predict another property. The same approach applied by model developers during model development and assessment of performance of the model should be applied, unless properly justified. In case the (Q)SAR model relies on many physicochemical descriptors, and it is unfeasible to evaluate the reliability of each input, the focus should be on the most influential descriptor(s).

- Each principle is broken down to assessment elements (AEs)
- AEs are further explained in the Guidance and Checklist
- The Guidance also explains the conditions for acceptable predictions

Figure: Guidance text with explanation of the AEs for assessing QSAR Predictions Principle 1: a correct input

ECHA
EUROPEAN CHEMICALS AGENCY

Prediction Checklist



Prediction 1

when more than one prediction is considered, add a comment here to identify to which prediciton the checklist refers to (e.g. model name and/or predicted structure)

Principle	Assessment element	Weight	Outcome	Uncertainty Comments
		Default values		Only for elements that are fulfilled
Correct input(s) to the model				
1.1	Clear and complete description of the input and model settings	High		
1.2*	Input representative of the substance under analysis	High		
1.3	Reliable input (parameters)	Medium		
Substance within the applicability domain of a valid model				
2.1	Substance within the applicability domain	High		
2.2	Any other limitation of the model is considered	High		
Reliable prediction				
3.1	Reproducibility	High		
3.2	Overall performance of the model	Medium		
3.3	Relationship of the substance with the physicochemical, structural and response spaces of the training set of the model	Medium		
3.4	Performance of the model for similar substances	High		
3.5*	Mechanistic and/or metabolic considerations	High		
3.6*	Consistency of information	High		
Outcome is fit for the regulatory purpose				
4.1*	Compliance with additional requirements	High		
4.2*	Correspondence between predicted property and property required by the regulation	High		
4.3*	Decidability within the specific framework	High		
Conclusion on the individual prediciton				
Uncertainty				
Outcome of the assessment (individual prediction)				
Comments				

For each assessment element (AE):

→ **Weight** - how important is the AE in the context of use of the prediction. It depends on the purpose of use of the prediction

- Low; Medium; High

→ **Outcome:**

- Fulfilled; Not fulfilled; Not applicable/assessed; Not documented

→ **Uncertainty -** how confident is the assessor with the outcome

- Low; Medium; High

By default, high uncertainty to AEs that are not fulfilled or not documented

Prediction Checklist

Prediction 1				
when more than one prediction is considered, add a comment here to identify to which prediciton the checklist refers to (e.g. model name and/or predicted structure)				
Principle	**Assessment element**	**Weight**	**Outcome**	**Uncertainty Comments**
		Default values		Only for elements that are fulfilled
Correct input(s) to the model				
1.1	Clear and complete description of the input and model settings	High		
1.2*	Input representative of the substance under analysis	High		
1.3	Reliable input (parameters)	Medium		
Substance within the applicability domain of a valid model				
2.1	Substance within the applicability domain	High		
2.2	Any other limitation of the model is considered	High		
Reliable prediction				
3.1	Reproducibility	High		
3.2	Overall performance of the model	Medium		
3.3	Relationship of the substance with the physicochemical, structural and response spaces of the training set of the model	Medium		
3.4	Performance of the model for similar substances	High		
3.5*	Mechanistic and/or metabolic considerations	High		
3.6*	Consistency of information	High		
Outcome is fit for the regulatory purpose				
4.1*	Compliance with additional requirements	High		
4.2*	Correspondence between predicted property and property required by the regulation	High		
4.3*	Decidability within the specific framework	High		
Conclusion on the individual prediciton				
Uncertainty				
Outcome of the assessment (individual prediction)				
Comments				

Conclusion

→ **Uncertainty of the prediction**

- Low; medium; High

Based on the highest uncertainty of high weight AEs.

→ **Outcome of the assessment**

- Acceptable for the intended purpose;
- Not acceptable for the intended purpose;
- Documentation insufficient to decide on the acceptance for the intended purpose.

The document suggests to accept predictions with low or medium uncertainty

“Prediction Criteria and uncertainty” spreadsheet

→ Also for predictions and results, a separate spreadsheet of the Checklist provides details, practical advice, examples and mapping to the QPRF for each AE

→ In addition, there is a section dedicated to how to assign the uncertainty level

Principle	Practical advice	Examples	Uncertainty		Mapping to mos
			This table offers guidance on how to assign the uncertainty level of each assessment element. *To assign the uncertainty for elements that are fulfilled, refer to the explanation in the column.* *For elements that are not fulfilled or not documented, high uncertainty should be assigned by default unless a valid justification is provided.* *For elements that are not applicable/assessed, leave empty* *NOTE: some examples include numeric values to explain more concretely how to proceed with the assessment . However, acceptable values depend on the predicted property and purpose of use of the prediction. The values used as examples should not be intended as thresholds established by the project.*		
Correct input(s) to			**Exaplanation of the uncertainty level**	**Examples**	
1.1	If the input is incomplete but the assessors are still able to reproduce the prediction, then the weight of this element in the overall assessment is lower.	Example 1: in case the model accepts as input the structure in form of SMILES, it is not sufficient to indicate as input the substance name and/or its numerical identifiers (such as CAS or EC numbers). Names and numerical identifiers may not unequivocally identify the SMILES that has been used as input. The exact SMILES used as input needs to be specified. Example 2: in case the model accepts as input three-dimensional structures, it is not sufficient to indicate as input the SMILES of the structure. Information on the three-dimensional structure, such a .mol file or equivalent, is needed.	**Low:** Input structure(s) and model settings are fully described **Medium:** some minor aspects of the input structure(s) and model settings are not clearly described **High:** some important aspects of the input structure(s) and model settings are not clearly described	A model requires SMILES and optionally logKow as input to generate a prediction. Low: SMILES and logKow provided Medium: SMILES provided, logKow not provided High: only CAS number provided, but CAS/SMILES association is ambiguous. **NOTE:** the reliability of logKow is assessed under AE 1.3	5 Input (all fields
1.2	The comparison can be done using expert judgment or by using publicly available information and tools that associate structures with names or other identifiers. If the model distinguishes the different tautomeric forms and generates different predictions, then it is important to indicate which form was used as input and justify the selection. If different tautomeric forms are investigated and produce the same prediction, this should also be indicated. If the model documentation indicates how to pre-process the input structure, possibly including how to represent tautomeric groups, these indications should be followed. Alternatively, the user should (if possible) use as input the structure in the tautomeric form that would be predominant if the corresponding experimental test were performed to measure the property of interest. Another option is to predict different forms and to calculate either a reasonable worst-case or an average, eventually weighted according to the abundance of the different forms.	Example 1: the substance under analysis is "formhaldeyde". The SMILES "C=O" is used as input. Using available resources, the correspondence between the name and the SMILES is verified. Example 2: the substance under analyis is a salt formed by an inorganic cation and an organic anion. The model does not accept the SMILES that includes both ions. The model documentation indicates that for salts, only the neutralised organic part should be used as input. The assessment consists in checking that the correct pre-processing has been followed. *Example 3 (for multiple predictions): the substance is formed by two major constituents. If two separate predictions are provided for the constituents, then the assessment element is fulfilled	**Low:** the composition of the substance under analysis is well covered by the input structure(s) **Medium:** the composition of the substance under analysis is mostly covered by the input structure(s) **High:** some constituents of the substance under analysis are not covered by the input structure(s)	The prediction refers to a substance that includes three constituents (one major constituent, one minor constituent and one impurity) in its composition. Low: predictions for all three constituents are provided Medim: predictions for two constituents are provided, impurity not considered High: only the prediction for the major constituent is provided	5 Input (all fields 2 Substance (all
1.3	Parameters that are automatically calculated by the model or software do not need to be evaluated at this stage.	An aquatic toxicity prediction is obtained from a model based on logKow. The prediction is generated by using as input an logKow defined by the user. The reliability of the user defined logKow needs to be verified.	**Low:** the values of the additional input parameters are associated with low uncertainty **Medium:** the values of additional input parameters are associated with medium uncertainty **High:** the values of additional input parameters are associated with high uncertainty	A model that requires manual input of logKow is used to generate a prediction. Low: the logKow value used as input is the result of a reliable experimental study Medium: the logKow value used as input is predicted by a QSAR model. No details are provided to assess its reliability. High: the logKow value used as input is predicted by a QSAR model. The prediction is unreliable, but it is the only available estimate.	5.2 Descriptors

Details and examples

Correct input – Assessment Elements (AEs)

→ AE 1.1: Clear and complete description of the input and model settings
- All information (input structure and/or parameters, model settings) is available to the assessors, thus making the prediction reproducible

→ AE 1.2: Input representative of the substance under analysis
- The structure(s) modelled represent the substance subject to regulatory assessment

→ AE 1.3: Reliable input (parameters)
- Parameters that are input manually (other than the chemical structure) are reliable

Correct input – example of assessment

→ AE 1.1: Clear and complete description of the input and model settings

What to check and how:

- It is clear whether the structure is input by using SMILES or other identifiers. If other parameters are also used as input, they are described

- If relevant, conformational (tri-dimensional) information is also given.

- In case of editable options, check if default settings are applied and, if not, if a justification is provided.

Example

A model requires SMILES and optionally logKow as input to generate a prediction.

Assessment:

→ Is the AE fulfilled? If yes, assign uncertainty:

- **Low** uncertainty: SMILES and logKow provided
- **Medium** uncertainty: SMILES provided, logKow not provided
- **High** uncertainty: only CAS number provided, but CAS/SMILES association is ambiguous.

Substance within the applicability domain of a valid model – AEs

→ AE 2.1: Substance within the applicability domain

- The substance meets the applicability domain (AD) requirements specified by model developers

→ AE 2.2: Any other limitation of the model is considered

- The substance does not meet any of the criteria for which the model should not be used

Applicability domain – example of assessment

→ AE 2.1: Substance within the applicability domain

What to check and how:

- For models that automatically calculate the AD, check that the substance is within AD

- When the AD is not calculated automatically, manually perform the AD assessment against the criteria specified by the developers.

Example

A model that automatically assesses the applicability domain is used.

Assessment:

→ Is the AE fulfilled? If yes, assign uncertainty:

- **Low** uncertainty: the model indicates that the substance is 100% within domain, and a clear explanation supports the claim
- **Medium** uncertainty: the model indicates that the substance is 100% within domain, but it is unclear how this is calculated
- **High** uncertainty: the model indicates that the substance is mostly within domain but some fragments of the substance are unknown to the model, therefore the substance cannot be considered to be fully within applicability domain

Reliable prediction – AEs

→ AE 3.1 Reproducibility
- The prediction can be reproduced using the same input and model version

→ AE 3.2 Overall performance of the model
- The model has an overall performance that is considered acceptable for the intended regulatory application

→ AE 3.3 Fit within the physicochemical, structural and response spaces of the training set of the model
- The prediction is result of interpolation in terms of physicochemical, structural and response space

→ AE 3.4 Performance of the model for similar substances
- The model predicts accurately substances similar to the one under analysis

→ AE 3.5 Mechanistic and/or metabolic considerations
- Mechanistic and metabolic considerations support the prediction

→ AE 3.6 Consistency of information
- Additional relevant and reliable information supports the prediction

Reliable prediction – example of assessment

→ AE 3.4: Performance of the model for similar substances

What to check and how:

- Check if the model predicts well substances similar to the one under analysis.

Example:

The predicted substance is a linear aliphatic saturated C8 secondary amine.

Assessment:

→ Is the AE fulfilled? If yes, assign uncertainty:

- **Low** uncertainty: data for other linear aliphatic saturated C6-C10 secondary amines are available, and the model predicts them well
- **Medium** uncertainty: data for other linear aliphatic saturated C3-C6 secondary amines are available, and the model predicts them well
- **High** uncertainty: data for other linear aliphatic saturated C6-C10 secondary amine are available, and the model predicts them fairly (one substance is misclassified by the model)

Outcome is fit for the regulatory purpose – AEs

→ AE 4.1: Compliance with additional requirements
- Regulation specific requirements for the use of computational results are met

→ AE 4.2: Correspondence between predicted property and property required by the regulation
- The modelled property corresponds to the property required by the regulation

→ AE 4.3: Decidability within the specific framework
- The outcome allows to take a regulatory decision in the framework of use

Reliable prediction – example of assessment

→ AE 4.2: Correspondence between predicted property and property required by the regulation

What to check and how:

- Check that the modelled property corresponds to the property required by the regulation

Example

The regulation requires the LC50 from a fish acute toxicity test according to OECD TG 203.

Assessment:

→ Is the AE fulfilled? If yes, assign uncertainty:

- **Low** uncertainty: the model predicts the LC50 from a fish acute toxicity test according to OECD TG 203
- **Medium** uncertainty: the model predicts the LC50 from a fish acute toxicity test after 96 hours. Other details such as fish species considered are not specified.
- **High** uncertainty: the predicted property is fish acute toxicity, no other details are specified.

Assessment of results based on multiple predictions

(Q)SAR results based on multiple predictions

Cases that consider multiple predictions include:

→ Predictions from different models for the same structure;

→ Predictions from the same models for different structures (such as the multiple constituents of a substance or for the substance under analysis and its metabolites);

→ A combination of the above.

Assessment workflow for results from multiple predictions

1. Within the Result Checklist, complete a checklist for each prediction individually (for complex cases, start by addressing multiple predictions associated with the same structure, and then consider the predictions for different structures)
2. Assess the additional AE:
   - Correct determination of the final result from individual predictions
3. Determine the uncertainty of the final result by weighing the uncertainty of individual predictions (e.g. consistent independent predictions lower uncertainty)
4. Decide on the acceptability of the result (the document suggests to accept results with low or medium uncertainty)

Determination of the final result – AE and example

→ AE 5.1: Correct determination of the final result from individual predictions

- Individual predicted values are aggregated correctly to determine the final result

What to check and how:

- Check that the (statistical) method used to determine the final result is explained

- If the regulation recommends specific rules (e.g. worst case approach), check that these are followed

Example:

The regulation requires a conservative approach when considering multiple reliable predictions.

Assessment:

Is the AE fulfilled? If yes, assign uncertainty:

Low uncertainty: two predictions are considered reliable and consistently predict low toxicity. The final result is low toxicity justified as consensus result.

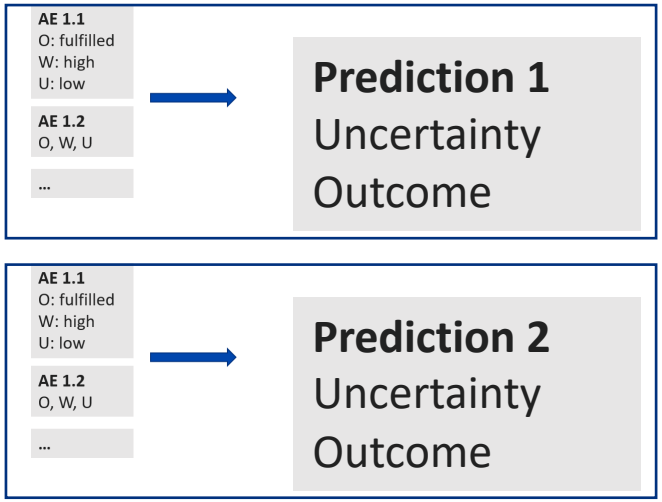
Medium uncertainty: two predictions are considered reliable and but produce slightly different results. One of the two values is preferred without justification.

High uncertainty: two predictions produce significantly different results. An average value is used as final result without justification.

Workflow for assessing results from multiple predictions

Assessment element (AE)
Outcome (O): fulfilled, not fulfilled, not documented, not applicable
Weight (W): low, medium, high
Uncertainty (U): low, medium, high
Conclusion: results acceptable, not acceptable, insufficient documentation

1. Assess predictions individually



2. Check how the final result is determined (AE 5.1)

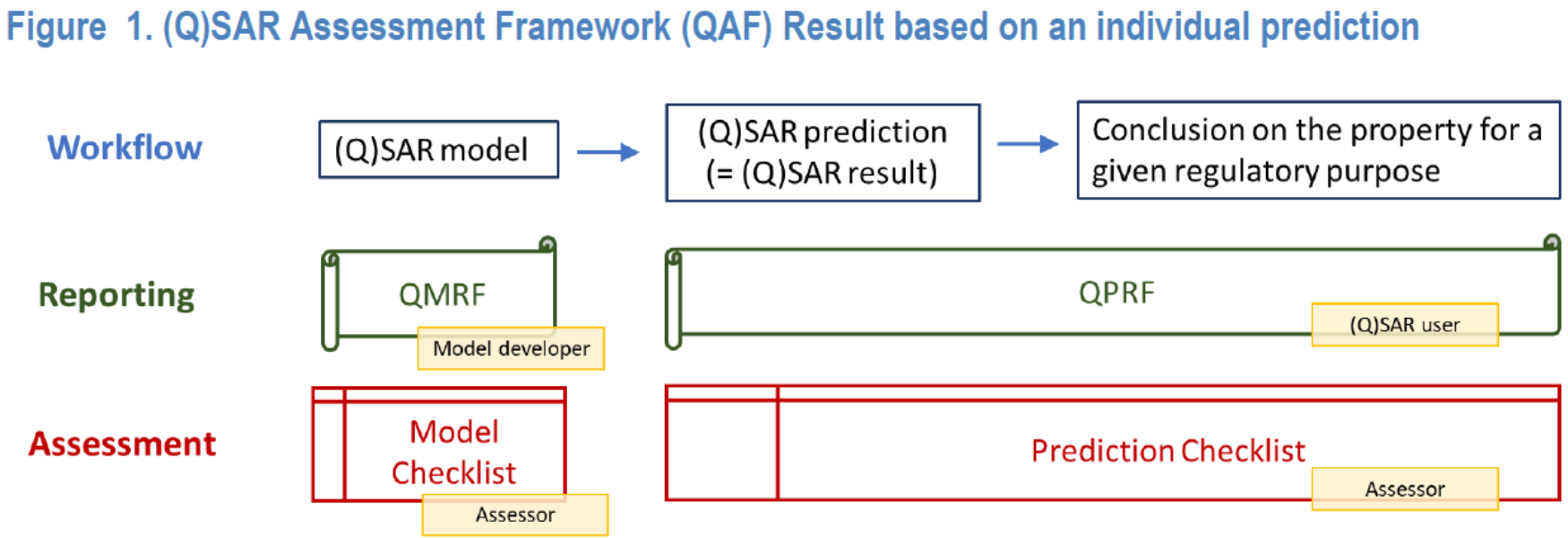
(Q)SAR result

3. Conclusion based on the level of uncertainty and purpose of use

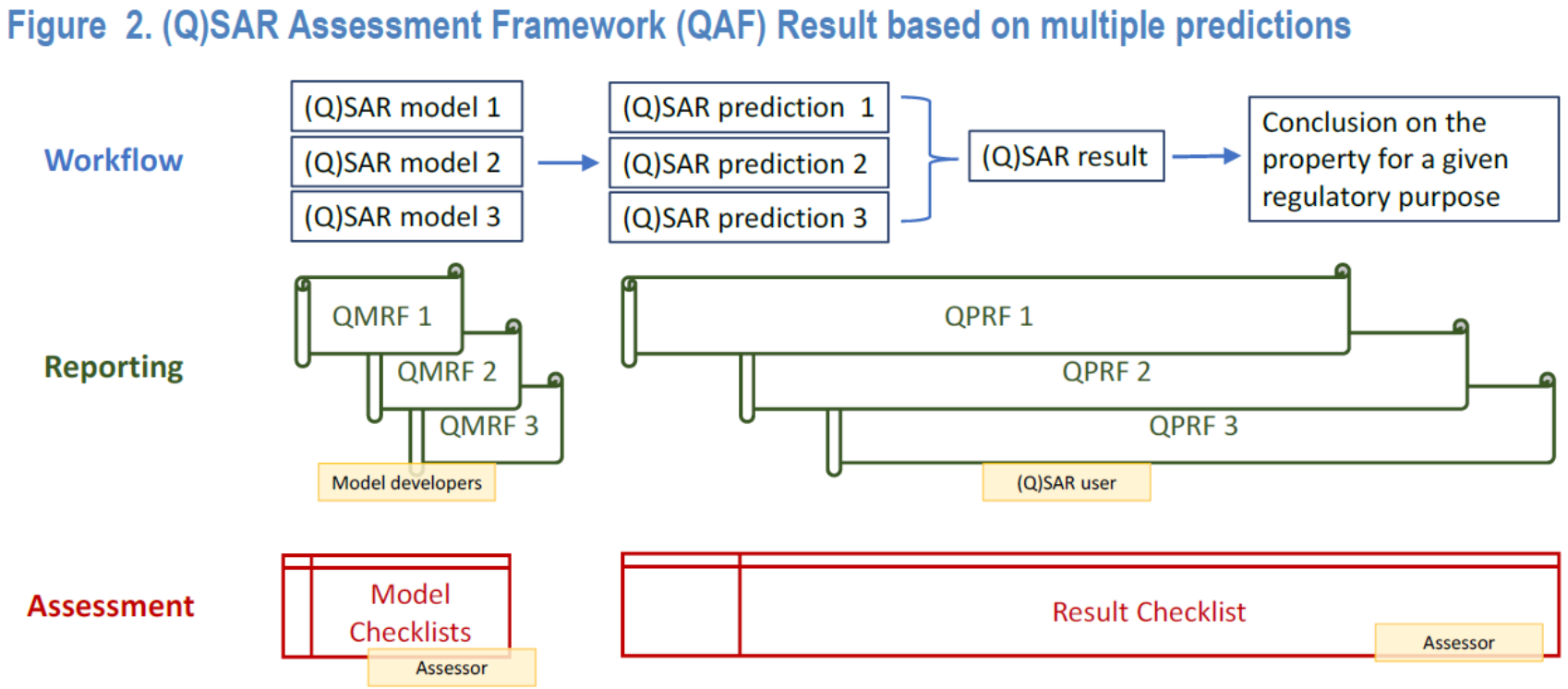
Conclusion on the result
Uncertainty
Outcome

ECHA
EUROPEAN CHEMICALS AGENCY

Visual abstract 1/2



Visual abstract 2/2



QAF Annexes – Updated QPRF and QMRF

Annexes:

- Updated **QSAR Prediction Reporting Format (QPRF v2.0)**: Major update to reflect the QSAR Assessment Framework Guidance. 8 main sections:

    1. General information
    2. Substance
    3. Model and software
    4. Prediction
    5. Input
    6. Applicability domain and limitations
    7. Reliability assessment
    8. Purpose of use (for regulatory applications)

- Updated **QSAR Model Reporting Format (QMRF v2.1)**: minor update because the OECD principles for the validity of models have not been changed

Conclusions

What is next

→ The OECD QAF expert group identified the following areas for further work:

- **Endpoint specific case studies** can be proposed under OECD IATA Case Study Project
- **Reporting** (extension of OECD Harmonised Templates to report QSAR information; a new report for results from multiple predictions)
- **Other** (update of the QMRF, technical annex on "external predictivity" of QSAR models)

Acknowledgments

QAF Contributors

→ **Co-leads from ISS**
- Olga Tcheremenskaia
- Cecilia Bossa
- Chiara Battistelli

→ **Co-lead from ECHA**
- Andrea Gissi

→ **OECD secretariat**
- Patience Browne
- Tomoko Aoyagi

→ **QAF Group Members**

QAF Group members*

Jessica Barkas (US Environmental Protection Agency (EPA))
Tara S. BARTON-MACLAREN (Health Canada)
Chiara Laura BATTISTELLI (National Institute of Health (ISS))
Michael BEKING (Environment and Climate Change Canada)
Mark BONNELL (Environment and Climate Change Canada)
Cecilia BOSSA (National Institute of Health (ISS))
Edoardo CARNESECCHI (European Food Safety Authority (EFSA))
Daniel Chang (US Environmental Protection Agency (EPA))
PAULA ECATERINA CHIRA (Ministry of Environment, Waters and Forests)
Anna Cruz (Australian Industrial Chemicals Introduction Scheme (AICIS))
Joop de Knecht (National Institute for Public Health and the Environment (RIVM))
Ayako FURUHAMA (National Institute of Health Sciences (NIHS))
Andrea Gissi (European Chemicals Agency (ECHA))
Marius Gudbrandsen (Climate and Pollution Agency)
Jörgen Henriksson (Swedish Chemicals Agency)
Matthias HERZLER (German Federal Institute for Risk Assessment (BfR))
Rune HJORTH (Danish Environmental Protection Agency)
Tina Hofmaier (Austrian Agency of Health and Food Safety (AGES))
Masashi HORIE (Ministry of Economy, Trade and Industry)
Miriam JACOBS (UK Health Security Agency)
Nobheetha Jayasekara (Australian Industrial Chemicals Introduction Scheme (AICIS))
Shinano KAWAHARA (Ministry of Economy Trade and Industry (METI))
Jeannette Koenig (Federal Institute for Risk Assessment (BfR))
Sunil KULKARNI (Health Canada)
Knud Ladegaard Pedersen (Danish Environment Protection Agency)
Donna Macmillan (Humane Society International (HSI))
Uko Maran (University of Tartu)
Julia Havilland MARKS (Department of Agriculture, Water and the Environment)
Todd Martin (US Environmental Protection Agency (EPA))
Enrico MOMBELLI (Institut National de l'Environnement Industriel et des Risques (INERIS))
Nikolai Georgiev NIKOLOV (National Food Institute)
Audrey Pearson (Environment Agency)
Martin Phillips (U.S. Environmental Protection Agency (EPA))
Prachi PRADEEP (Federal Institute for Risk Assessment (BfR))
Qinghong PU (Department of the Environment)
Sarah Robinson (Australian Industrial Chemicals Introduction Scheme (AICIS))
Emiel Rorije (National Institute of Public Health and the Environment (RIVM))
Yuki SAKURATANI (National Institute of Technology and Evaluation)
Gerrit SCHÜÜRMANN (UFZ Centre for Environmental Research)
Rositsa Serafimova (European Food Safety Authority (EFSA))
Yongkwon SONG (Permanent Delegation of Korea to the OECD)
Olga TCHEREMENSKAIA (Istituto Superiore di Sanita (ISS))
David Tobias (EPA-Office of Pollution Prevention and Toxics)
Manuela Trobej (Austrian Agency of Health and Food Safety (AGES))
Petra Van Kesteren (National Institute for Public Health and the Environment (RIVM))
Eva Bay WEDEBYE (National Food Institute)
Andrew WORTH (European Commission (Joint Research Centre))
Takashi YAMADA (National Institute of Health Sciences (NIHS), Division of Risk Assessment, Center for Biological Safety and Research)

*official list from OECD

Thank you

andrea.gissi@echa.europa.eu

echa.europa.eu/subscribe

Connect with us

echa.europa.eu/podcasts

European Chemicals Agency

@one_healthenv_eu

@EU_ECHA

@EUECHA

EUchemicals