Webinar: OECD QSAR Assessment Framework in REACH dossier evaluation: what you need to know

Questions and answers

ECHA organised a webinar on 21 March 2024 on the OECD QSAR Assessment Framework in REACH dossier evaluation.

This document is based on the questions received before and during the webinar. Editorial changes have been made to improve clarity and similar questions have been combined.

For the most up-to-date advice on restrictions, contact us or refer to our support material.

#	Question	Answer
Reporting, IUCLID and registrations		
1	Thank you very much for organizing this webinar. I would like to know if registrants are expected to attach models and predictions checklists to IUCLID?	No, registrants are not expected to compile or attach the QAF checklist to the dossiers. Regulators will eventually complete the checklist. Nevertheless, registrants may go through the checklist for their QSAR studies to ensure that they are valid, and all documentation has been provided.
2	How should the checklists be filled for UVCBs or multi-constituent substances when there are many constituents?	For multi-constituents and UVCB, we expect registrants to provide one prediction for each constituent and then take all predictions into account to conclude on the property of the substance. Regulators will use the QAF Result checklist to assess the validity of

#	Question	Answer
		the approach.
3	When we update a REACH registration dossier, are we supposed to revise QSAR predictions and fill in the additional fields in IUCLID, or is it enough to do it for new dossiers / endpoint study records? When do QSARS used in registration dossiers need to be updated and why? Similarly, when we make a QSAR today, is it necessary to update this QSAR in 6months or 1 year time from today?	Whenever you are updating your registration dossier, you should check whether all the information which you have provided before is up to date and relevant. If you think that your existing QSAR study is compliant according to the QAF, you do not need to update it. For now, if you did not fill in the additional new QSAR fields, especially those which Andrea mentions in his presentation, there are no additional requirements or technical completeness checks that would lead to rejection of your dossier. Those fields will be introduced in the next release of IUCLID to be more explicit on certain elements.
4	How to report in IUCLID QSAR results based on multiple predictions? Is there a plan for a new IUCLID ESR (similar to the WoE justification) for summarizing multiple predictions in a QSAR result? If multiple models are used to obtain a result, do all of the models QMRFs and QPRFs need to be provided in the IUCLID or should there be a joint QMRF and QPRF for one result?	As of today, the best way to report a QSAR result is to fill in an endpoint study record for each prediction, and then one more record to report the reasoning on how the different predictions are put together to get the result. However, we know that there is the need to have a harmonised QSAR result reporting format to give the possibility to structure the information on QSAR results. ECHA is leading together with EFSA a project to develop such result reporting format. Based on this reporting format, the same project will also give advice on how this could be further implemented in IUCLID. Right now we are looking at the possibility to have individual endpoint study record for each prediction and then one additional endpoint study record for the QSAR result, but this is work in progress, since the concept of QSAR result in QAF is new.
5	How do you weight QSAR predictions, and the QSAR result, in Dossier Evaluation when you have multiple tests outcomes and results? Do you use only the QSAR result?	The question is interpreted as follows: I have multiple QSAR predictions, they may even disagree among each other, and then I have a QSAR result: what is ECHA going to check, the QSAR result or each of the QSAR predictions? We start by looking at the QSAR result, but to verify that the QSAR result was correctly determined we will also have to go back to the individual predictions, so we are going to check everything. We expect that the QSAR result justification explains why the conclusion goes into a certain direction. Maybe one of two predictions was unreliable and out of the domain, so only the other one is taken into account.
6	When are you going to start using QAF for compliance checks?	It was explained during the webinar. Assessment Framework from our practice point of view is not the novelty. We have encoded with the OECD expert group our experience from 10 years of evaluating QSAR results within REACH in the QSAR assessment framework. Our practices are not really changing. What may change in the future is how we are referring to the different legal arguments in our decisions. This will happen from the next year. If we will start to refer to the Assessment Framework principle in our decisions, then we will update ECHA guidance on QSAR and grouping, and we will make the reference for the QAF there to have consistent legal basis for our decision.
7	Are the new IUCLID fields for the QSARS parts of the TCC already now or when are they implemented in REACH IT as business rule	The new QSAR fields that we will be available in April in IUCLID will not be part of the technical completeness check, at least not now. In future we may change this, but we are not going to do anything that is too heavy from registrants' side. As an example, in

#	Question	Answer
	failure?	future we may make mandatory completing the field of applicability domain, which is a simple picklist.
8	Do all available models for a specific endpoint need to be used and then combined to a result (provided they are applicable and give reliable results)? When is it enough to use one specific model to fill a data gap?	No, there is no need to use all models. One valid prediction can be enough. However, it may be in the interest of the registrant to check more than one model and to see if their predictions are in agreement. This is because if there are discrepancies, it is better for the registrant to understand why and if one of the models does not predict well a substance property. Still, if the chemistry or the endpoint requires a more cautious approach it is advised to address the property with more than one model.
9	Is it still requested to provide a RSS for the respective endpoint for each of the RA substances used in the QSAR result?	Robust study summaries are not needed when analysing analogues to assess QSARs. However, the information on the studies must be good enough to consider the experimental data as reliable.
10	What is the correct approach when no QMRF is provided by the model developer or available online? Would that be a knock-out criterium to use the prediction in general?	If a model lacks QMRF, we suggest using a different model, at least for REACH purposes. For most endpoints there are many models available with adequate documentation. If the information is provided but in a different format than QMRF, that is no problem.
11	Do the output files from the QSAR model need to be submitted and are these sufficient to fulfil requirements to submit a QPRF?	It is good practice to provide the raw output file from the QSAR software. However, when the output file does not contain certain elements from the QPRF, it is not sufficient and a QPRF must be provided.
12	The slide of the new fields that will be added to the next release of IUCLID, showed fields to add similar substances with experimental data. If the substance is part of a category, will it be necessary to duplicate this information here?	The information on analogues in the new IUCLID fields is very limited, so only a small part of the information will eventually need to be repeated.
Guidance and examples		
13	Could you provide a list of QSARs for which no QMRF is needed (as the information is publicly available)?	No, we cannot provide such list. As an example, a QMRF could be available online at the time the dossier is prepared, but not available anymore when the dossier is evaluated by regulators. For this reason, you may want to always attach the QMRF to the dossiers. In any case, if QMRF is not provided, we check for its availability online to proceed with the assessment.
14	When QRMFs are assessed by authorities in the context of a Registration, will the results be shared publicly? And is there an OECD effort to consolidate the different authority's experts views?	There is no plan to share the results of assessments of QMRF. The challenge with sharing assessments of QMRF is that the assessments depend on the purpose of the use of the model and its predictions. Therefore, OECD cannot publish a list or consolidate different expert views from authorities because they can legitimately differ if the purpose of use of the model is different. Furthermore, we have seen in the past misunderstandings that if a model is considered acceptable, then there was the expectation by the registrants that all its predictions will be valid too.
15	Will you provide best practice examples QAFs for the most frequently used QSAR Models? If not, may you explain why you came to this	OECD has published a few examples prepared by regulators on the use of the QAF checklist, which were also reviewed by ECHA. However, there are no examples for REACH prepared by ECHA. We are right now in the process of finalising our internal

#	Question	Answer
	decision? Is it possible for ECHA to produce some detailed case studies where step by step QSAR model is developed and the outcome is very robust e.g at least 200 similar compounds in the dataset, R2 is at least above 0.75 and the full dataset (just SMILES and say IC50 values in a csv file)?	policies and workflows to use the QSAR assessment framework in our assessment. One year from now, when we start to use the framework in compliance check, we could be in the position to prepare some examples.
16	More practical examples for common models (Biowin,ECOSAR,...) would be helpful or are such examples already available? If we have already the examples, are they still applicable for the QAF as well?	Some examples are presented in ECHA Guidance R7, ECHA practical guide on the use of QSAR and previous webinars. We have not yet published any example updated according to the QAF.
17	These seminars are very insightful - thank you. Is it possible to produce some detailed webinars on specific QSAR models - e.g., how they were developed, the dataset selected and also including the statistics (e.g. high R" values) and comparing the predictions to actual experimental results?	One thing that we are going to experiment from next month is to invite model developer to give a seminar for ECHA and if everything goes fine we are also going to make this seminar publicly available. There are a number of challenges in doing that, one thing is that if the model is commercial then it looks like we are endorsing or giving a commercial advantage to a commercial developer, so we prefer to start with seminars on freely available tools. We do not want to create expectations that if something is presented in an ECHA seminar it is automatically compliant and will be accepted, so we will have disclaimers about it. If everything goes fine, at least for some public tools we may be able to invite the developers to give such information. Also keep in mind that many developers publish articles and present at conferences. So if it does not come from ECHA, you may be able to find this information somewhere else. There were already quite a few webinars available on many tools, including practical examples on how you can assess the reliability of the predictions.
18	Will you update ECHA Guidance on QSARs (R6) to reflect the publication of the QAF?	We are not planning to update this guidance at the moment. We may eventually add an introductory note with a link to the QAF to inform that we may refer to it in our decisions. We do not see the need for substantial update of the guidance, which we believe is aligned with the QAF.
19	How are you going to assess QSAR predictions already submitted to ECHA? According to R.6 or QAF?	For now, we will keep referring to ECHA Guidance R6, which is anyway in accordance with the QAF.
20	A list of acceptable QSAR models would be helpful. Does ECHA plan to publish such a list? Can ECHA provide a list of QSAR models that meet or almost meet the QAF and R.6 guidelines? It may be helpful to have this list which will help the developers as feedback and help them provide missing info (QRMF) to	Not for now, because the list will need to be comprehensive and maintained, and this is difficult in a rapidly evolving field.

#	Question	Answer
	achieve full compliance. This would help promote the use of QSARs for NAM.	
21	Please comment on the use of global and / or local models in terms of diverse datasets for predicting toxicity in QSAR models. Can both be used? Can you post some links where these were used ideally in a specific dossier and / or in a ECHA publication / video?	Both local and global models can be used. If both produce valid predictions for a substance of interest, we recommend submitting both predictions.
Specific assessment elements		
22	Can you confirm that being outside the applicability domain for certain descriptors does not mean that the model is unacceptable, if it can be discussed as acceptable considering the other descriptors that define the domain for a prediction model?	The short answer is yes, we can confirm that. The whole point of the QAF is to list what is important, but also to leave then the flexibility to assessors to focus on what matters. When there is a parametric range and the substance is just a little outside this range, this substance would be formally out of applicability domain. On the other hand, if we can see that there are very good analogues in the training set that are well predicted by the model and everything else seems to be fine in the prediction, we are not going to fail the prediction only because formally it happens that one parameter is outside the applicability domain. However, if we see other issues in addition to the applicability domain, most likely we are going to list all the issues that we found, and among these issues that we have found we may also pick on the fact that formally the substance is outside the applicability domain.
23	Is there any plan for harmonizing the AD definition of models? Or at least to set some "minimum requirements" for AD definition?	Short answer no, at least not by ECHA. This was discussed during the QAF meeting of the expert group. The challenge we had there was that the QAF was a framework discussed among regulators, while for the definition of applicability domain we needed the counterpart of model developers. However, there has not been agreement in many years on how to best define applicability domain and we are satisfied of the pragmatic solution that was found in the QAF. We are going to use whatever definition of applicability domain the model developers give. On the other hand, we know that there are some aspects that are very important to assess the validity of a prediction irrespective of what is the definition of applicability domain given by the model developers, and all these aspects is what we have gathered under the reliability of the prediction.
24	What is the level of performance of a QSAR model ECHA is expecting to accept a QSAR prediction?	ECHA does not have a legal reference to reject QSAR studies in dossier evaluation solely based on poor performance of the model. However, predictions from poor performing models often present other issues that are then picked up and communicated by ECHA in its decisions. In future, ECHA is considering referring to the R2 and Q2 values given in the QAF Checklist as examples as minimum requirements.
25	Expert models, such as Derek Nexus do not have a clearly defined applicability domain, so	Also negative predictions need to fulfil QAF requirements. Lack of alerts in absence of other information to perform a validity assessment according to the QAF are not acceptable. We note that some software distinguish predictions for lack of alerts and

#	Question	Answer
	how reliable are negative predictions in those models?	negative predictions, where the latter could be used for regulatory purposes if compliant with validity requirements for QSAR results.
26	What kind of weight will ECHA assign to the Assessment Elements for predictions? Will it be those default ones indicated in the checklists? Is the same weight assigned to assessment elements for all endpoints, or will there be differences?	Information equivalent to all AEs in the QAF is currently assessed in our evaluation activities. This means that we give high weight to all AE, and we do not have a priori distinction depending on the endpoints. Nevertheless, we have some flexibility. We do not expect that a prediction is "perfect". We aim not to reject any predictions which are good enough.
27	Could you explain AE 3.2 once more? To my understanding this is already assessed in the QMRF why assessing this once more in the prediction?…	When assessing a model (QMRF), we check that the statistical measures of performance of the model are provided. In the context of a specific prediction, we check that the performance is good enough for the purpose of use.
Similarity and analogues		
28	Similarity is a "broad" concept. Can be calculated with different methods and considering different structural features. Any plan for harmonization or for giving advices? How to identify analogues – different methods for similarity, different analogues?	There is no need to focus on the mathematical methodology to find analogues, as we do not use the number provided by calculations of similarity with different methodologies. What we do is that we look with our expertise on the endpoints if the analogues make sense, if they can really provide information on how good the prediction is for the new substance. If from the endpoint point of view the analogues make sense, they are acceptable, irrespective of the tanimoto coefficient (or other algorithms used for similarity). Please do not focus on the mathematical methodology, but focus on the endpoint, mechanism and what make sense for an expert perspective to inform on the local performance of the model.
29	Can you please elaborate how you determine that the model predicts well with similar substances?	It depends on the purpose of use of the prediction. As an example, if the property needs to be assessed against a regulatory threshold (e.g. BCF = 2000 L/kg), and the similar substances are all below the threshold and accurately predicted, then it is very good indication that there is no concern. If the predicted similar substances have experimental and predicted values close to both sides of the regulatory threshold and the prediction for the substance is quite close to the threshold, then the assessment is more challenging. Overall, the information from similar substances should give the regulators confidence that the model performs well in this local chemical space.
30	If I want to show that my QSAR works well for similar substances I most probably have to refer to experimental data (eg dissiminated data on the ECHA website) - do I need any rights to do so?	No, to demonstrate good local performance of a QSAR model there is no need to have the right to use the "source" data. This is because the analysis of local performance of a QSAR model only uses the part of the results that is free from intellectual property (i.e. the result value and metadata around it).
31	Will 3D similarity be acceptable or just 2D similarity based models?	3D models are acceptable, but we do not see many in REACH dossiers.
Data and tools		

#	Question	Answer
32	What kind of data for models development do you consider reliable? Do data must be obtained from OECD TGs? If yes, it may limit the possibilities to develop new acceptable models and it would slow the shift to non-animal approaches?	Ideally, when using a model for a prediction under REACH registration, every point used to develop the model (i.e. in the training set) should be of sufficient quality, ideally generated following the OECD Test Guideline required under REACH to cover information on the target property/endpoint. In real cases, for most QSAR models we do not have sufficient information to assess the quality of each data point individually, or we may know that not all data points meet the required quality standards. We often have a general description of the data curation procedure, which is difficult to verify. To avoid rejecting most of the predictions submitted, in these cases we look at the performance of the model and data quality for the closest analogues. If the performance on close analogues, based on good quality data, is satisfactory, we may accept the prediction despite a "not perfect" training set. Furthermore, if all data points in the training set follow a trend, and this trend also include some data for which the data quality is unclear, the fact that the points follow the trend increases the confidence in the use of these data.
33	How availability of data may affect the external validation and is it mandatory for regulatory acceptance (e.g. in cases the software is proprietary, and training sets are not available) + Protected data in commercial software?	External validation is used to demonstrate that a model is not overfitted on its training set. Overfitting may lead to inaccurate predictions for new structures. Overfitting is more likely to occur with higher number of descriptors and the use of advanced statistical techniques that are designed to achieve high R2 values without necessarily improving external predictivity. With this premise in mind, external validation is more important for statistical models (where good statistics are the main tool to evaluate the validity of the model) rather than mechanistic models. When using a mechanistic model, a mechanistic explanation of the result and good local performance in presence of relevant analogues has more weight than external validation in the assessment of the validity of a prediction.
34	Hello, is the framework going to accept the results of DANISH QSAR?	There are two different platforms: the Danish QSAR Database and the Danish QSAR Models. The Danish QSAR Database is a database of predictions and in terms of applicability domain and reliability it only includes information whether the prediction is in domain or not, without explaining why for the specific case. There are no means to evaluate independently the applicability domain and reliability of that prediction, so we usually do not accept predictions from the Danish QSAR Database. It is different for the Danish QSAR Models because these models provide results with more details and can be evaluated according to the REACH requirements.
35	With regard to the use of an unambiguous algorithm many commercial software programs use their own algorithms and are not explained in detail, e.g., such as using Pymol or autodock. Is it enough to show what docking procedures were undertaken via "known" methodology but algorithm is not included?	We do not see many docking models in REACH. For us the point is that somehow we must be able to understand what is happening. The challenge I could see in the case described in the question is that we may not have access to the tool and to the algorithm, so we would have to "trust" the result that is given to us without being able to reproduce it. That is going to be difficult if it ever happens. Ideally, it would be good when the algorithm is not disclosed to at least make the tool available to enable us to

#	Question	Answer
		reproduce the prediction to see that what is being reported is not just a result that is convenient for the assessment, but that it is really the result that is given by the model.
36	When or is it ever acceptable to make a qualitative assessment, when the result is of a highest or lowest extreme?	It depends on the purpose. If the purpose is a qualitative assessment, then a qualitative outcome is be good enough. However, often the value is needed also for risk assessment, for which a quantitative output may be required. In this latter case, a qualitative value will not be sufficient.
37	How to deal with commercial tools, where a lot of information is "hidden" to the user, due to confidentiality?	Results from commercial tools can be used for REACH purposes when they come with sufficient information to assess the validity of the model and of the prediction.
OECD QSAR Toolbox		
38	Will the QSAR toolbox create the QPRF in the new format including the new sections? ECHA develops the OECD QSAR Toolbox. Is it compliant with QAF?	Yes, it will. The next Toolbox public release is planned for June, and will include the updated QPRF and other features related to the QAF. We note that the QSAR Toolbox is often used under REACH for read across predictions, and as clarified in separate ECHA webinars, in this case the study and justification should refer to ECHA read-across assessment framework (RAAF).
39	Do all QSARs in the ECHA/OECD QSAR Toolbox have the QMRFv2.1 as background and thus could be used under REACH? Also do these models have the QPRFv2.0 already?	QSAR Toolbox will be updated with the new formats in the next release in June. Other models will probably be updated too in due time. We note that even predictions reported with older versions of QMRF and QPRF can be considered compliant (also according to the QAF) as long as they fulfil the criteria in the Checklist.
40	How does ECHA plan to evaluate QMRF reports generated from OECD QSAR toolbox for completely new molecules?	Please refer to ECHA webinar: New developments and regulatory applications of the OECD QSAR Toolbox (https://echa.europa.eu/-/new-developments-and-regulatory-applications-of-the-qsar-toolbox).
41	QSAR Toolbox – the models seem to be based on one descriptor at a time such as just logP, implying that just one descriptor is enough (along with similar structures. How is that the regulators are not requiring the use of multiple descriptors for use for example, multiple regression analysis?	Simple models are sometimes preferred to more complex ones. They are easier to interpret and less likely to be overfitted. Predictions from models with one descriptor are subject to the same requirements as predictions from more complex models.
Endpoint specific		
42	Biodegradation simulation studies on UVCBs or multi-constituent substances are technically very challenging. But QSAR models that produce actual half-lives and metabolite formation rates are scarce. What combination of (Q)SAR models and types of output could be accepted to cover this endpoint? Since biodegradation is a very complex environmental procedure, how justified is the use of QSAR	There are good models for ready biodegradability. However, for simulation studies we are not aware of models that would produce acceptable half-life predictions. In case of UVCB or degradation products, these models can be used for screening and refining testing strategies.

#	Question	Answer
	models for prediction of biodegradation rate of substances or mixture?	
43	Are mass Balance models for bioaccumulation also accepted when a model description is provided?	Mass balance models address the fate of a chemical. They are not classical QSAR models. Mass balance models are kinetic models and can give valuable insights. Sometimes they are used in combination with QSARs, for example as input. As long as the model addresses the relationship between a structure and a property, e.g. structure and bioconcentration factor, we will try to assess the model using the assessment elements of the QAF.
Artificial Intelligence (AI)		
44	What is the ECHA view on current developments in using AI models to predict chemical properties? Is it acceptable (now, nowadays), to use NN and AI models for regulatory purposes (i.e. when the algorithm is not available)?	We start to see more and more QSAR models based on AI. We have already seen and assessed them in the last years, and in some cases we have accepted their results in compliance check. In practice, there is no difference in our assessment whether the model predictions come from AI based model or more traditional models. As long as the documentation is clear, we can follow the logic on how the prediction has been derived and we can assess the local performance of the model, this is acceptable for us. In terms of unambiguous algorithm, if we can run the tool and reproduce the prediction, that's fine for us.
45	Thank you for this webinar. AI is very much in the news. Are there any plans to review and assess the potential use of AI in predicting toxicity?	In principle when we are assessing the predictions we will not discriminate whether the model prediction is from AI based model or a "traditional" model. When you are building the model you have always two elements, first is how to compile extensive robust and reliable training and validation sets, and then the second part is how to associate different structural features to predict the properties and to demonstrate that you have a correlation. AI can help with both elements, because we can see more and more that AI algorithms look for new data, can already pre-assess to some extent quality of these data and start to build almost kind of simultaneously the training and validation sets. When we are talking about finding the relationship between different structural features and the predicted property then this non-parametric approach which AI offers may be as well very powerful and it might give you the good results and we probably will see more and more those models in the future. However, as I said before, as long as it is well explained how the data were identified, what were the curation protocol, what kind of quality assurance the developers took into account to make sure that the data which were used to build the model was of good quality and what kind of algorithm were used to find this relationship between properties and the structural features then it is fine. Anyway, you have to check how the model performs in the local chemical space, how the close analogues are predicted, whether the close analogues are close enough. From our perspective there is no big difference. It is also a matter of how AI is defined. There is a lot of discussion on what is AI. A potential issue is the data quality behind the model. So if a model searches the internet for data, we do not know the quality of the data. So this is then for us to assess. We need to be sure that the input for the model is

#	Question	Answer
		adequate for our purpose and only then we can use the output, meaning the prediction. Therefore, it is very important to report how quality is assured because you can use those AI algorithms to identify additional data point, but then at least you need to apply clear filtering criteria (quality criteria) on how you basically narrow down those information and make them available for the model. If by AI you are thinking about chatGPT, do not submit us the chatGPT prompts whether the substance is toxic for reproduction. First because most probably chatGPT will hallucinate a bit about it, second because tomorrow you may get completely different answers. So the idea about reproducibility is completely gone. If your question was whether you can use chatGPT then my answer is not yet.
46	ChatGPT and other LLMs are going to become very important in regulatory approvals. Can a chatGPT produced QSAR model still be included in a dossier submission if the data it produces is also validated in another program, eg QSAR toolbox / excel)?	We are not there yet. We will consider this possibility in due time. In general, the predictions will have to fulfil the same requirements as "traditional" QSARs.