



Organización para la Cooperación y el Desarrollo Económicos (OCDE)

ENV/CBC/MONO(2023)32

DIRECCIÓN DE MEDIO AMBIENTE COMITÉ DE SUSTANCIAS QUÍMICAS Y BIOTECNOLOGÍA

Marco de Evaluación de (Q)SAR: Guía para la evaluación regulatoria de modelos de Relación Estructura-Actividad (Q)SAR, predicciones y resultados basados en múltiples predicciones

Serie sobre Pruebas y Evaluación

N° 386

Dos documentos anexos están disponibles con las siguientes referencias:

- ENV/CBC/MONO(2023)32/ANN1
- ENV/CBC/MONO(2023)32/ANN2

La lista de verificación (versión EXCEL) está disponible en el siguiente enlace:

<https://www.oecd.org/chemicalsafety/testing/QAF-Checklist.xlsx>

Contenido	
Prólogo.....	3
Resumen ejecutivo.....	4
Resúmenes visuales.....	6
1 Evaluación de Modelos de (Q)SAR (Lista de Verificación de Modelos)	7
1.1 Endpoint definido ²	7
1.2 Algoritmo no ambiguo	8
1.3 Un dominio de aplicabilidad definido	9
1.4 Medidas apropiadas de bondad de ajuste, robustez y predictividad	9
1.5 Interpretación mecanística.....	10
1.6 Resultado de la evaluación del modelo	11
1.7 Conclusiones sobre la evaluación del modelo.....	11
2 Evaluación de Predicciones de (Q)SAR (Lista de Verificación de Predicciones).....	12
2.1 Entrada(s) correcta(s) al modelo.....	12
2.2 Sustancia dentro del dominio de aplicabilidad	13
2.3 Fiabilidad de la(s) predicción(es).....	14
2.4 Adecuación del resultado para el propósito regulatorio.....	16
2.5. Conclusión sobre la evaluación de una predicción individual.	17
3 Evaluación de un Resultado de (Q)SAR derivado de múltiples predicciones (Lista de Verificación de Resultados)	18
3.1 Cuándo utilizar la Lista de Verificación de Resultados.....	19
3.2 Incertidumbre y resultado de la predicción (Q)SAR.....	19
4 Consideraciones Finales.....	21

Prólogo

En noviembre de 2004, en la 37ª Reunión Conjunta del Comité de Sustancias Químicas de la OCDE y el Grupo de Trabajo sobre Sustancias Químicas, Plaguicidas y Biotecnología (Reunión Conjunta), se acordaron los "Principios de la OCDE para la Validación, con Fines Regulatorios, de Modelos (Q)SAR".

En 2007, el mismo Grupo de Trabajo de la OCDE publicó el "Documento de orientación sobre la validación de modelos de Relación Cuantitativa entre la Estructura y la Actividad [(Q)SAR]" para el uso de (Q)SAR en aplicaciones regulatorias (OCDE, 2007).

Desde entonces, la posibilidad de utilizar (Q)SAR se ha incluido en diversas regulaciones relacionadas con sustancias químicas. Esto ha permitido que diferentes partes interesadas adquieran experiencia práctica en la evaluación de (Q)SAR para diferentes fines regulatorios.

Esta experiencia ha resaltado que no todas las predicciones producidas por un modelo válido son aceptables para todos los fines regulatorios. Cuando una predicción (Q)SAR o un resultado generado a partir de múltiples predicciones se utiliza para un determinado fin regulatorio, es necesario verificarlo en el contexto de la aplicación específica. Aunque hubo acuerdo en los principios para la evaluación de modelos, era necesario establecer un conjunto de principios comúnmente acordados para la evaluación regulatoria de predicciones (Q)SAR y resultados de múltiples predicciones.

A finales de 2020, se propuso al Grupo de Trabajo de la OCDE sobre Evaluación de Peligros (WPHA) un proyecto dirigido por el Istituto Superiore di Sanità (ISS) para desarrollar un Marco de Evaluación (Q)SAR (QAF), y se añadió al plan de trabajo a principios de 2021. El proyecto incluye los principios para evaluar modelos y predicciones (Q)SAR en forma de una lista de verificación y criterios para evaluar los elementos de la lista. Además, la Agencia Europea de Sustancias y Mezclas Químicas (ECHA) se unió al ISS como co-líder del proyecto.

Se envió una solicitud de nominación de expertos a los delegados del Grupo de Trabajo de Evaluación de Peligros (WPHA) y se convocó a un grupo de más de 40 expertos para proporcionar aportes y revisar el QAF. El Grupo de Expertos del QAF se reunió a través de una serie de teleconferencias entre 2021 y 2023, y subgrupos más pequeños contribuyeron a la redacción. Una reunión presencial del Grupo de Expertos del QAF se llevó a cabo en noviembre de 2022 para ayudar a finalizar el documento de borrador.

Con este documento, el grupo de expertos del QAF de la OCDE ha establecido principios de la OCDE para la evaluación de predicciones (Q)SAR y resultados basados en múltiples predicciones, y acordó listas de verificación para realizar la evaluación de modelos, predicciones y resultados de múltiples predicciones en la práctica.

Este documento se publica bajo la responsabilidad del Comité de Sustancias Químicas y Biotecnología de la OCDE.

Resumen ejecutivo

El objetivo del Marco de Evaluación de Relaciones Cuantitativas entre la Estructura y la Actividad ((Q)SAR) es desarrollar un marco sistemático y armonizado para la evaluación regulatoria de modelos (Q)SAR, predicciones y resultados basados en múltiples predicciones. La evaluación propuesta pretende ser aplicable independientemente de la técnica de modelado utilizada para construir el modelo, el endpoint predicho y el propósito regulatorio previsto. El público principal de este documento son las autoridades regulatorias y sus partes interesadas. Además, se alienta a cualquier otro usuario de (Q)SAR a consultar el QAF cuando utilice (Q)SAR con fines regulatorios.

La evaluación de (Q)SAR para fines regulatorios no debe limitarse a verificar la validez del modelo utilizado, ya que incluso un modelo válido puede producir predicciones inaceptables en ciertas condiciones. Por lo tanto, las predicciones individuales y los resultados de múltiples predicciones requieren evaluaciones específicas. Con este propósito, el QAF se basa en los principios de la OCDE para la validación de modelos (OCDE, 2007, referidos como Principios de Modelos (Q)SAR de la OCDE en el resto del documento) y define nuevos principios para la evaluación de predicciones (Q)SAR y resultados basados en múltiples predicciones. Se han establecido cuatro principios (referidos como Principios de Predicción (Q)SAR de la OCDE en el resto del documento) relacionados con 1) la entrada correcta, 2) la adecuación de la sustancia dentro del dominio de aplicabilidad del modelo, 3) la confiabilidad de la predicción y 4) la idoneidad del resultado para el propósito identificado. Para simplificar la evaluación, cada principio se ha subdividido en elementos que deben tenerse en cuenta en la evaluación (Elementos de Evaluación, EEs). Los EEs se incluyen en tres listas de verificación (Lista de Verificación del Modelo, Lista de Verificación de Predicciones y Lista de Verificación de Resultados, proporcionadas en un documento aparte) que se pueden utilizar para evaluar la aceptabilidad del uso de (Q)SAR en la práctica. Cada EE puede cumplirse o no, no estar documentado o no ser aplicable. Las Listas de Verificación también proporcionan más detalles y ejemplos para cada EE.

La Lista de Verificación del Modelo consta de una lista de EEs para evaluar un modelo de acuerdo con los Principios de Modelos (Q)SAR de la OCDE (OCDE, 2007). La Lista de Verificación del Modelo debe usarse junto con las otras Listas de Verificación al evaluar predicciones y resultados basados en múltiples predicciones. En este caso, el uso de un modelo aceptable se puede considerar el primer paso de la evaluación. Si el modelo se considera aceptable, entonces la evaluación debe considerar las otras Listas de Verificación. Cuando un modelo se considera inaceptable, la evaluación podría concluir sin considerar más las predicciones y los resultados. Alternativamente, la Lista de Verificación del Modelo se puede usar como una herramienta independiente cuando, por ejemplo, se utilizan (Q)SAR para analizar bases de datos sin la posibilidad de evaluar las predicciones individualmente, o para mantener un registro separado de la evaluación de un modelo que podría reutilizarse en el futuro. La evaluación de un modelo es específica para el propósito regulatorio y debe repetirse al evaluar el mismo modelo para otros propósitos.

Las listas de Verificación de Predicciones y Resultados se utilizan para evaluar las predicciones individuales (Q)SAR y los resultados basados en múltiples predicciones, respectivamente. Estas listas de verificación están compuestas por EEs basados en los Principios de Predicción (Q)SAR de la OCDE, que tienen diferentes pesos según su importancia para la evaluación. Se proporcionan valores predeterminados sugeridos para los pesos en las Listas de Verificación de Predicciones y Resultados, pero los evaluadores pueden modificar su importancia para que se ajusten a su propio marco regulatorio o paradigma. Además, los evaluadores pueden asignar un valor de incertidumbre semicuantitativo (bajo, medio, alto) a cada elemento de evaluación, siguiendo esta guía y los ejemplos en las Listas de Verificación. Finalmente, la incertidumbre general de la predicción se determina considerando la incertidumbre asociada con cada EE y su peso en la evaluación. Según el propósito de uso y el nivel de incertidumbre de la predicción individual, el evaluador puede concluir sobre el resultado de la evaluación (es decir, si la predicción es aceptable para el propósito previsto de uso).

En el caso de resultados basados en múltiples predicciones, además de la evaluación individual de cada predicción, la Lista de Verificación de Resultados considera un EE adicional para evaluar si las predicciones se integran correctamente para determinar el resultado final. El nivel de incertidumbre del resultado final se asigna ponderando la incertidumbre de las predicciones individuales y el EE adicional. Finalmente, también en este caso, el evaluador concluye sobre el resultado de la evaluación según el propósito de uso del resultado, lo que determina el nivel de incertidumbre aceptable (es decir, si el resultado es aceptable para el propósito previsto de uso).

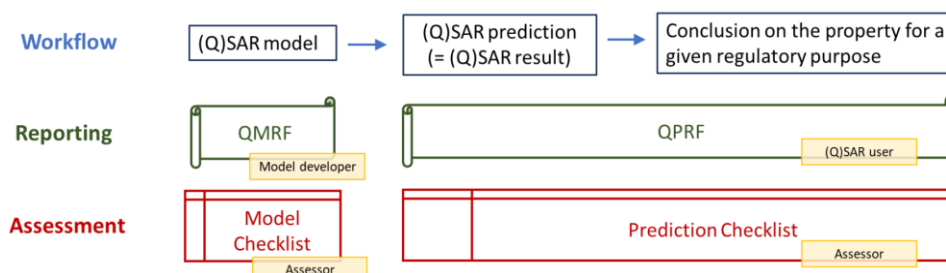
Como complemento al QAF, se han desarrollado actualizaciones para el formato de informe de modelos (Q)SAR (QMRF) y el formato de informe de predicciones (Q)SAR (QPRF). Mientras que la actualización del QMRF se refiere únicamente a la descripción de la información esperada en cada campo, sin cambios en los nombres y el orden de los campos, el QPRF se actualizó de manera más extensa para reflejar los Principios de Predicción (Q)SAR de la OCDE recién establecidos. Las plantillas actualizadas de QMRF y QPRF se proporcionan como anexos de este documento.

Además, se han desarrollado ejemplos que ilustran el uso de las Listas de Verificación y se proporcionarán como documentos separados. El grupo de expertos recomienda el uso y la aplicación continuos de los principios del QAF en el marco del proyecto de estudios de casos IATA de la OCDE para los estudios de casos que incluyan enfoques (Q)SAR.

Por último, algunos expertos en el grupo han identificado la necesidad de una guía adicional sobre cómo medir la predictividad externa de los modelos (Q)SAR, y se inició una discusión hacia el final del proyecto QAF. Sin embargo, el tema no se desarrolló en la versión final del QAF debido a que no hay consenso científico sobre cómo medir la predictividad externa, y por lo tanto, este tema requeriría una revisión exhaustiva de la literatura que está fuera del alcance de este proyecto. Además, la validación de modelos es una tarea para los desarrolladores de modelos y, como tal, está fuera del alcance del QAF, que tiene como objetivo principal proporcionar instrucciones para los reguladores que evalúan modelos preexistentes.

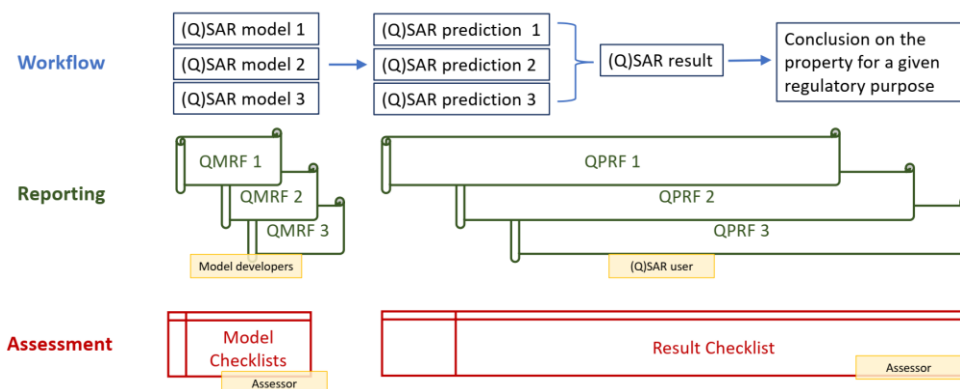
Resúmenes visuales

Figura 1. Marco de Evaluación (Q)SAR (QAF) Resultado basado en una predicción individual



Nota: Flujo de trabajo de información de (Q)SAR para un resultado basado en una predicción individual según el marco de evaluación (QAF) de la OCDE para (Q)SAR. La información sobre el modelo se reporta en el Formato de Informe de Modelos (QMRP) de (Q)SAR, preparado por el desarrollador del modelo y evaluado por los reguladores mediante la Lista de Verificación de Modelos del QAF. La información sobre la predicción de (Q)SAR se reporta en el Formato de Informe de Predicciones (QPRF) de (Q)SAR por el usuario de (Q)SAR y es evaluada por los reguladores utilizando la Lista de Verificación de Predicciones del QAF. Las listas de verificación también podrían ser precompiladas por el usuario de (Q)SAR para facilitar el trabajo del evaluador.

Figura 2. Marco de Evaluación (Q)SAR (QAF) Resultado basado en múltiples predicciones



Nota: Flujo de trabajo de información de (Q)SAR para un resultado basado en múltiples predicciones según el marco de evaluación (QAF) de la OCDE para (Q)SAR. La información sobre los modelos se reporta en los documentos de Formato de Informe de Modelos (QMRP) de (Q)SAR, preparados por los desarrolladores de modelos y evaluados por los reguladores mediante la Lista de Verificación de Modelos del QAF. La información sobre las predicciones y el resultado de (Q)SAR se reporta en el documento de Formato de Informe de Predicciones (QPRF) de (Q)SAR por el usuario de (Q)SAR, y es evaluada por los reguladores mediante la Lista de Verificación de Resultados del QAF.

Nota: El Formato de Informe de Predicciones (QPRF) está diseñado para informar predicciones individuales y tiene campos limitados para describir la integración de múltiples predicciones en un resultado individual. Por esta razón, la publicación de una plantilla adicional para informar resultados de (Q)SAR basados en múltiples predicciones se está considerando.

1 Evaluación de Modelos de (Q)SAR (Lista de Verificación de Modelos)

Esta sección del documento proporciona consejos prácticos para completar la Lista de Verificación de Modelos. Se basa en y complementa la orientación más teórica de la OCDE sobre la validación de modelos (OCDE, 2007).

Los Principios de la OCDE para la validación de (Q)SAR establecen que "para facilitar la consideración de un modelo de (Q)SAR con fines regulatorios, debería estar asociado con la siguiente información:

1. un endpoint definido
2. un algoritmo no ambiguo
3. un dominio de aplicabilidad definido
4. medidas apropiadas de bondad de ajuste, robustez y predictividad
5. una interpretación mecanística, si es posible".

Los capítulos siguientes y la Lista de Verificación de Modelos proporcionan más detalles sobre estos principios y cómo verificar que un modelo de (Q)SAR cumple con ellos. La evaluación debe basarse en la información proporcionada en el Formato de Informe de Modelos de (Q)SAR (QMRF). También se proporciona un mapa de los EEs en la Lista de Verificación de Modelos a los campos del QMRF en la Lista de Verificación de Modelos para facilitar la recuperación de la información relevante para la evaluación. Cada uno de los Principios de Validación de (Q)SAR de la OCDE también se considera más a fondo en uno o más EEs de las Listas de Verificación de Predicciones y Resultados, como se indica al final de cada subcapítulo a continuación.

1.1 Endpoint definido²

Según el Principio 1 de la OCDE para Modelos de (Q)SAR (OCDE, 2007), un (Q)SAR debe estar asociado con un "endpoint definido", donde endpoint se refiere a cualquier propiedad fisicoquímica, biológica o ambiental que pueda medirse y, por lo tanto, modelarse. La intención de este principio es asegurar la transparencia en el punto final que está siendo predicho por un modelo dado, ya que un punto final podría determinarse mediante diferentes protocolos experimentales y bajo diferentes condiciones experimentales.

La Lista de Verificación de Modelos incluye los siguientes Elementos de Evaluación (EEs) para verificar que el punto final esté claramente definido:

1. Propósitos científicos y regulatorios claros
2. Transparencia de los datos experimentales subyacentes
3. Calidad de los datos experimentales subyacentes

Propósitos científicos y regulatorios claros (EE 1.1 en la Lista de Verificación del Modelo)

Para tener un propósito científico claro, la propiedad predicha debe describirse de manera precisa. Para tener un propósito regulatorio claro, un modelo debe abordar un requisito regulatorio específico, que a menudo está asociado con un método de prueba o una guía de prueba específica, o debe proporcionar información de apoyo a dicho requisito (por ejemplo, información mecanicista). La descripción de la propiedad predicha debe ser lo más detallada posible, incluyendo todos los elementos que se han considerado (por ejemplo, la unidad de medida, escala de tiempo, observaciones como crecimiento, mortalidad, etc.). La complejidad de la propiedad predicha influye en el alcance de la documentación requerida (es decir, los modelos que predicen propiedades más complejas, como la toxicidad del desarrollo, requieren más detalles en la definición de la propiedad en comparación con los modelos que predicen propiedades más simples, como la mutagenicidad in vitro en el ensayo de Ames).

Transparencia de los datos experimentales subyacentes (EE 1.2 en la Lista de Verificación del Modelo)

Este EE se refiere a la transparencia de los datos experimentales subyacentes y del procedimiento de selección y curación de datos relacionados. Las fuentes de los datos experimentales deben ser informadas de manera adecuada, al igual que la información sobre los criterios de selección de datos experimentales, el procesamiento de datos y la información sobre identificadores químicos (incluido al menos un identificador que codifique la estructura química, como InChi/InChiKey o SMILES (canónico), y otra información comúnmente reportada como números de registro CAS) de las sustancias probadas.

También se deben investigar posibles sesgos en la selección de datos (por ejemplo, la inclusión sistemática en el conjunto de entrenamiento de datos medidos de acuerdo con guías de prueba no relacionadas con el endpoint predicho). Los estudios originales (o una referencia accesible) representan el nivel más alto de transparencia, pero rara vez están disponibles. Por el contrario, los estudios subyacentes pueden no estar disponibles en absoluto para algunos modelos debido a la confidencialidad o documentación insuficiente. Para muchos modelos (Q)SAR existentes, el nivel de transparencia se encuentra entre estos dos extremos, con algunos, pero no todos los detalles disponibles para los estudios experimentales utilizados para construir los modelos.

Las autoridades responsables de la evaluación pueden decidir el nivel mínimo aceptable de transparencia necesario para propósitos específicos, teniendo en cuenta que, para algunos modelos, la información disponible podría ser limitada por razones comerciales, por ejemplo. En general, debería haber información suficiente sobre los datos subyacentes o sobre el procedimiento de curación de datos para poder evaluar la calidad de los datos.

Calidad de los datos experimentales subyacentes (EE 1.3 en la Lista de Verificación del Modelo)

El modelo (Q)SAR debe construirse sobre datos de calidad suficiente. Sin embargo, la evaluación individual de la calidad de cada punto de datos a menudo no es factible. En estos casos, la calidad de los datos subyacentes puede evaluarse en función de la descripción del procedimiento de curación de datos. Por ejemplo, los evaluadores pueden verificar cómo se han tenido en cuenta los parámetros experimentales relevantes (por ejemplo, sexo, especie, temperatura, período de exposición, protocolo) que podrían afectar los resultados de los estudios experimentales al seleccionar datos para construir el modelo. Los evaluadores también pueden considerar si todos los puntos de datos utilizados para desarrollar y validar un modelo son generados mediante el uso de 1) el mismo protocolo de ensayo; y 2) el protocolo de ensayo más actualizado, y cuáles son las consecuencias para la confiabilidad. La calidad de los datos individuales también debe evaluarse en la medida de lo posible.

El Principio de Validación 1 de la OCDE (Q)SAR se considera además en las Listas de Verificación de Predicción y Resultados bajo el elemento "Correspondencia entre la propiedad predicha y la propiedad requerida por la regulación".

1.2 Algoritmo no ambiguo

De acuerdo con el Principio 2 del Modelo (Q)SAR de la OCDE (OCDE, 2007), un modelo (Q)SAR debe estar expresado en forma de un algoritmo no ambiguo (entendido como una descripción inequívoca del algoritmo). El propósito de este principio es asegurar la transparencia en la descripción del algoritmo del modelo para permitir la reproducibilidad independiente de sus predicciones. La Lista de Verificación del Modelo incluye los siguientes Elementos de Evaluación (EE) para verificar el principio de un algoritmo inequívoco:

1. Descripción del algoritmo y/o software
2. Entradas y otras opciones
3. Accesibilidad del modelo

Descripción del algoritmo y/o software (EE 2.1 en la Lista de Verificación del Modelo)

El primer elemento a verificar es la disponibilidad de una descripción transparente del algoritmo. La ecuación del modelo, si corresponde, incluyendo todos los descriptores y el enfoque utilizado para su selección, debe ser detallada. Además, si corresponde, se debe proporcionar una lista de fragmentos/alertas de estructura (por ejemplo, activos, inactivos, máscaras) y su descripción. También se puede incluir la justificación que guió su identificación.

Los descriptores calculados deben ser señalados con el nombre y la versión del software utilizados para su cálculo. Además, se debe proporcionar la versión, la información de contacto de los desarrolladores y cualquier descripción disponible del software para el modelo (Q)SAR. Cuando una descripción exacta del algoritmo no esté disponible públicamente (por ejemplo, en el caso de modelos comerciales), cualquier información relevante disponible aún debe ser evaluada.

Entradas y otras opciones (EE 2.2 en la Lista de Verificación del Modelo)

En segundo lugar, los evaluadores deben verificar si la documentación incluye una descripción de las entradas y configuraciones del software del modelo. Se debe documentar los formatos de entrada permitidos (o preferidos) para la estructura química y sus descriptores, incluidos los procedimientos de preprocesamiento aplicables (por ejemplo, para sales y tautómeros). Además, se deben informar y explicar las opciones/configuraciones personalizables en el software. A

menos que se justifique lo contrario, se espera que los formatos de entrada y opciones recomendados sean los mismos que los utilizados por los desarrolladores del modelo al desarrollar el modelo y evaluar su rendimiento.

Accesibilidad del modelo (EE 2.3 en la Lista de Verificación del Modelo)

Finalmente, se debe verificar si la versión del modelo bajo evaluación es públicamente accesible. Se espera un enlace funcional para acceder o descargar el modelo en la documentación QMRF. Cuando los evaluadores tienen acceso a una versión diferente del modelo bajo evaluación (por ejemplo, una versión más reciente), se deben investigar las diferencias en los resultados.

El Principio 2 del Modelo (Q)SAR de la OCDE se considera además en las Listas de Verificación de Predicción y Resultados bajo el elemento "Reproducibilidad". Tenga en cuenta que cuando el modelo se implementa en un programa de software al que el evaluador puede acceder, la reproducibilidad de los resultados debería ser posible incluso en casos en que la descripción del algoritmo no esté completamente revelada. Los evaluadores pueden decidir que esto es aceptable para algunos propósitos regulatorios.

1.3 Un dominio de aplicabilidad definido

De acuerdo con el Principio 3 del Modelo (Q)SAR de la OCDE (OCDE, 2007), un modelo válido está asociado con un dominio de aplicabilidad definido (AD, por sus siglas en inglés). El dominio de aplicabilidad de un modelo (Q)SAR, como se describe en la guía (OCDE, 2007), es el espacio de respuesta y estructura química en el cual el modelo realiza predicciones con una confiabilidad determinada. Ampliando la definición de AD proporcionada anteriormente, el AD debe considerar el espacio paramétrico, estructural, mecanístico, metabólico y de respuesta del modelo. No obstante, el Marco de Evaluación de QSAR (QAF, por sus siglas en inglés) no prescribe una forma específica de definir el AD de un modelo, ya que pueden utilizarse múltiples metodologías válidas.

Estos se describen en la guía (OCDE, 2007), el cual puede consultarse para obtener más detalles sobre los aspectos científicos relacionados con el dominio de aplicabilidad (AD), mientras que este párrafo se centra en los aspectos prácticos de la evaluación dentro del Marco de Evaluación de QSAR (QAF, por sus siglas en inglés).

La Lista de Verificación del Modelo incluye un Elemento de Evaluación (EE) relacionado con el dominio de aplicabilidad:

- Definición clara del dominio de aplicabilidad y limitaciones del modelo

Definición clara del dominio de aplicabilidad y limitaciones del modelo (EE 3.1 en la Lista de Verificación del Modelo)

Los evaluadores deben verificar que la definición propuesta por los desarrolladores del modelo sea lo suficientemente detallada para permitir la evaluación de cómo una sustancia específica se relaciona con el dominio de aplicabilidad (AD) del modelo en las Listas de Verificación de Predicción y Resultados (por ejemplo, si la sustancia está dentro del AD del modelo, etc.). Para facilitar la evaluación, se alienta a los desarrolladores de nuevos modelos a implementar funcionalidades para incluir automáticamente en el informe de predicción información sobre cómo la sustancia de entrada se relaciona con el AD. Además, las Listas de Verificación de Predicción y Resultados incluyen una evaluación separada de confiabilidad que considera aspectos cruciales que pueden influir en la confiabilidad del resultado, independientemente de si están incluidos en la definición del AD, como se describe en el Capítulo 2.3 a continuación.

La documentación del modelo puede incluir una lista de limitaciones del modelo además de la definición del AD, como clases de sustancias para las cuales no se recomienda el uso del modelo. El QAF no es prescriptivo en términos de cómo definir tales limitaciones, pero debe verificarse que la descripción de limitaciones adicionales potenciales sea lo suficientemente detallada para evaluar cómo se relaciona una sustancia de entrada específica con ellas.

El Principio 3 se considera además en las Listas de Verificación de Predicción y Resultados bajo los elementos "Sustancia dentro del dominio de aplicabilidad" y "Se considera cualquier otra limitación del modelo".

1.4 Medidas apropiadas de bondad de ajuste, robustez y predictividad

De acuerdo con el Principio 4 (OCDE, 2007), un (Q)SAR debe estar asociado con "medidas apropiadas de bondad de ajuste, robustez y predictividad". Este principio expresa la necesidad de proporcionar información sobre la bondad de ajuste y robustez de un modelo (según se determina mediante validación interna) y la predictividad de un modelo (según se determina mediante validación externa). El dominio de aplicabilidad y el rendimiento de un modelo están conectados. El

rendimiento debe medirse dentro del dominio de aplicabilidad definido por sus desarrolladores. Por lo general, se puede lograr un mejor rendimiento del modelo con un dominio de aplicabilidad más estrecho. La guía (OCDE, 2007) puede consultarse para obtener más detalles sobre los aspectos científicos relacionados con el Principio 4.

La Lista de Verificación del Modelo incluye los siguientes Elementos de Evaluación (EE) para verificar la adecuación de las medidas de bondad de ajuste, robustez y predictividad del modelo:

- Bondad de ajuste, robustez
- Predictividad

Bondad de ajuste, robustez y predictividad (EE 4.1 y 4.2 en la Lista de Verificación del Modelo)

Los evaluadores deben verificar que la información disponible sobre el rendimiento interno y externo del modelo pueda usarse como una indicación de la precisión esperada del modelo al predecir nuevas sustancias. El tamaño de los conjuntos de entrenamiento, prueba y externos, los métodos y métricas estadísticas utilizados, los valores de las métricas estadísticas y la transparencia del procedimiento para medir el rendimiento son algunos de los aspectos importantes a considerar al evaluar el rendimiento del modelo. Los conjuntos de entrenamiento, prueba y externos deben ser independientes.

La información que describe cómo se seleccionó el modelo QSAR en cuestión y cómo se estimó su rendimiento predictivo debe ser evaluada para verificar si el modelo fue entrenado correctamente sin tener en cuenta la información que se utilizará para estimar la predictividad externa del modelo. Por ejemplo, el uso de información de un conjunto de pruebas externo de sustancias químicas (originalmente diseñado exclusivamente para la evaluación de la predictividad externa) durante el desarrollo del modelo (es decir, "fuga de datos" del conjunto de pruebas para la calibración o selección del modelo) es muy probable que resulte en una sobreestimación de la predictividad externa del modelo. El Principio 4 de la OECD para los modelos QSAR se considera aún más en las Listas de Verificación de Predicción y Resultado bajo el elemento "Rendimiento general del modelo".

1.5 Interpretación mecanística.

De acuerdo con el Principio 5 (OCDE, 2007), un (Q)SAR "debería estar asociado con una interpretación mecanística, si es posible". Los métodos estadísticos utilizados para describir las relaciones entre la estructura química y la actividad no pretenden reemplazar otros conocimientos de química y toxicología cuando existen dichos conocimientos. Los evaluadores pueden requerir que la documentación del modelo incluya consideraciones sobre cómo la justificación detrás de un modelo (Q)SAR es coherente o tiene en cuenta el conocimiento relacionado con la propiedad predicha (como las Rutas de Resultado Adverso conocidas, AOP, relevantes para la propiedad predicha), es decir, una interpretación mecanística. Las consideraciones toxicocinéticas también forman parte de la interpretación mecanística, si son relevantes para la propiedad de interés.

La Lista de Verificación del Modelo incluye el siguiente Elemento de Evaluación (EE) relacionado con la interpretación mecanística:

- Plausibilidad de la interpretación mecanística

Plausibilidad de la interpretación mecanística (EE 5.1 en la Lista de Verificación del Modelo)

Para modelos basados en fragmentos o alertas, la interpretación mecanística puede basarse en la explicación de la reactividad química o la interacción molecular causada o inhibida por los fragmentos asociados con las alertas. Para modelos basados en ecuaciones, la interpretación mecanística puede basarse en la interpretación fisicoquímica de cada descriptor y su asociación con un modo o mecanismo de acción. También se espera una indicación de si la base mecanística del modelo se determinó a priori (por ejemplo, preseleccionando descriptores o fragmentos para ajustarse a un mecanismo de acción específico) o a posteriori (por ejemplo, después del modelado, mediante la interpretación del conjunto final de estructuras y/o descriptores de entrenamiento).

El Principio 5 del Modelo (Q)SAR de la OCDE se considera además en las Listas de Verificación de Predicción y Resultados bajo el elemento "Consideraciones mecanísticas y/o metabólicas".

1.6 Resultado de la evaluación del modelo

La evaluación de cada Elemento de Evaluación (EE) y del modelo en general depende de (y es específica para) un propósito regulatorio dado.

Un modelo debe considerarse aceptable cuando el resultado de cada EE esté "cumplido", o alternatively, cuando haya suficiente información sobre el modelo disponible para evaluar los Elementos de Evaluación relacionados con el modelo en las Listas de Verificación de Predicción y Resultados. Para algunos propósitos regulatorios y con una justificación válida, también pueden aceptarse modelos que no cumplan con todos los Elementos de Evaluación.

1.7 Conclusiones sobre la evaluación del modelo

La compilación de la Lista de Verificación del Modelo es el primer paso en la evaluación de predicciones y resultados de múltiples predicciones. También respalda la evaluación de Elementos de Evaluación (EE) relacionados con la validez del modelo en las Listas de Verificación de Predicción y Resultados. Una vez completada, la Lista de Verificación del Modelo se puede reutilizar, es decir, los evaluadores no necesitan volver a evaluar el modelo cada vez que se envía una nueva predicción. Además, la Lista de Verificación del Modelo puede funcionar como una herramienta independiente para evaluar modelos utilizados para la búsqueda en bases de datos para las cuales las predicciones no están destinadas a evaluarse individualmente.

Finalmente, la Lista de Verificación del Modelo incluye una correspondencia entre la Lista de Verificación del Modelo y el formato del modelo (QMRP). Esta correspondencia, junto con el resultado de la evaluación de la Lista de Verificación del Modelo, cuando se divulga, puede servir como retroalimentación para los desarrolladores de modelos para una mejora adicional de sus modelos y la documentación relacionada.

2 Evaluación de Predicciones de (Q)SAR (Lista de Verificación de Predicciones)

Con el fin de evaluar las predicciones (Q)SAR y los resultados de múltiples predicciones, se necesitaban establecer principios para su evaluación regulatoria, además de los Principios del Modelo (Q)SAR. El (QAF) establece cuatro principios para la evaluación de predicciones (Q)SAR y resultados de múltiples predicciones con fines regulatorios:

1. las entradas del modelo deben ser correctas;
2. la sustancia debe estar dentro del dominio de aplicabilidad del modelo;
3. la(s) predicción(es) deben ser confiables;
4. el resultado debe ser adecuado para el propósito regulatorio.

Los mismos principios se incluyen en las Listas de Verificación de Predicción y Resultados. Este capítulo y las Listas de Verificación de Predicción y Resultados proporcionan más detalles sobre estos principios y cómo verificar que una predicción individual o un resultado de múltiples predicciones cumpla con ellos. La evaluación debe basarse en la información proporcionada en el formato del modelo (QMRF) y el Formato de Predicción y Resultados (QPRF). Este capítulo describe los Elementos de Evaluación (EE) comunes a las Listas de Verificación de Predicción y Resultados, mientras que el siguiente capítulo se centra en los elementos adicionales a considerar en la Lista de Verificación de Resultados.

2.1 Entrada(s) correcta(s) al modelo

Los modelos (Q)SAR requieren uno o más elementos de entrada para generar una predicción. Dependiendo del modelo, la entrada puede estar limitada a información sobre la estructura de la sustancia o también puede incluir algunos de sus descriptores. Algunos modelos también tienen configuraciones personalizables en el software.

La entrada se considera correcta cuando está claramente y completamente descrita, es representativa de la sustancia que se está analizando y utiliza parámetros confiables (por ejemplo, valores para descriptores que se usarán en la predicción). La documentación necesaria para asegurarse de que una entrada sea correcta depende de la complejidad del modelo (o del software que lo implementa) y de la sustancia bajo análisis. En general, la entrada debe prepararse siguiendo cuidadosamente las instrucciones de los desarrolladores del modelo o el software, si están disponibles.

La Lista de Verificación de Predicción incluye los siguientes Elementos de Evaluación (EE) para verificar que la entrada sea correcta:

- Descripción clara y completa de la entrada y configuraciones del modelo
- Entrada representativa de la sustancia bajo análisis
- Entrada confiable (parámetros)

Descripción clara y completa de la entrada y configuraciones del modelo (EE 1.1 en las Listas de Verificación de Predicción y Resultados)

El primer elemento a verificar es la descripción de la entrada y asegurarse de que sea inequívoca y completa. En el caso más simple, el modelo toma información sobre la estructura (por ejemplo, SMILES) como la única entrada y no tiene otras opciones editables que acompañen la entrada estructural. En este caso, la descripción de la información estructural exacta y la versión del modelo/software que se utilizaron para obtener la predicción son suficientes. Para casos más complejos, el requisito es proporcionar toda la información, incluida la información tridimensional sobre la estructura química, opciones personalizables ("configuraciones") y parámetros de la aplicación de software (por ejemplo, entrada manual de valores de los descriptores y su origen) que se necesitan como entrada para el modelo.

Entrada representativa de la sustancia bajo análisis (EE 1.2 en las Listas de Verificación de Predicción y Resultados)

En segundo lugar, es importante verificar que la entrada sea representativa de la sustancia bajo análisis y, por lo tanto, relevante para su evaluación. Cuando la sustancia consta de un único constituyente bien definido, es suficiente verificar la concordancia entre el nombre de la sustancia, la estructura y los identificadores numéricos. Para modelos tridimensionales, se espera información sobre la justificación para la selección de la conformación utilizada como entrada. Para sustancias con composiciones complejas, un resultado (Q)SAR puede derivarse de múltiples predicciones que cubran los constituyentes e impurezas. De hecho, una de las ventajas de los (Q)SAR es que se pueden predecir más constituyentes y metabolitos para investigar su contribución a la toxicidad general de la sustancia con costos adicionales limitados.

Además, algunos modelos pueden requerir que las entradas sean sometidas a una curación estructural antes de poder utilizarse para una predicción. Esto ocurre a menudo en casos de sales, estructuras ionizables o estructuras sujetas a tautomerismo, por ejemplo. En estos casos, existen diferentes enfoques. La elección del enfoque debe decidirse caso por caso y se debe prestar especial atención a cómo se realizó el preprocesamiento por parte de los desarrolladores del modelo para las sustancias del conjunto de entrenamiento y a las recomendaciones del marco regulatorio de interés, si son relevantes.

Entrada confiable (parámetros) (EE 1.3 en las Listas de Verificación de Predicción y Resultados)

Finalmente, para los modelos que utilizan entradas directas más allá de la estructura química, como un(des) descriptor(es) fisicoquímico(s), la fuente del valor de ese descriptor, ya sea medido experimentalmente o predicho por otro modelo, debe evaluarse para determinar su confiabilidad antes de usarlo para predecir otra propiedad. El mismo enfoque aplicado por los desarrolladores del modelo durante el desarrollo del modelo y la evaluación del rendimiento del mismo debe ser aplicado, a menos que esté debidamente justificado. En caso de que el modelo (Q)SAR se base en muchos descriptores fisicoquímicos y sea inviable evaluar la confiabilidad de cada entrada, el enfoque debe centrarse en el(des) descriptor(es) más influyente(s).

2.2 Sustancia dentro del dominio de aplicabilidad

El segundo principio requiere que el modelo sea aplicable a la sustancia bajo análisis. La evaluación de la aplicabilidad del modelo a la sustancia se basa en verificar cómo la sustancia bajo análisis se relaciona con el dominio de aplicabilidad (AD) y las limitaciones del modelo tal como las definen los desarrolladores del modelo. Si hay aspectos que potencialmente influyen en la confiabilidad de la predicción y que no han sido considerados por los desarrolladores del modelo al definir el dominio de aplicabilidad, estos pueden evaluarse al analizar el siguiente principio "Predicciones confiables". El dominio de aplicabilidad informa sobre la confiabilidad de la predicción. En este documento, el AD y la confiabilidad se evalúan por separado para agilizar el procedimiento de evaluación.

La Lista de Verificación de Predicción incluye los siguientes Elementos de Evaluación (EE) para verificar que la sustancia esté dentro del dominio de aplicabilidad del modelo:

- Sustancia dentro del dominio de aplicabilidad
- Consideración de cualquier otra limitación.

Sustancia dentro del dominio de aplicabilidad y consideración de cualquier otra limitación (EE 2.1 y 2.2 en las Listas de Verificación de Predicción y Resultados)

En general, la definición del AD debe incluir criterios para evaluar si la sustancia bajo análisis está dentro del AD. En algunos casos, los desarrolladores del modelo también proporcionan una descripción de las limitaciones conocidas del modelo y de los casos en que el modelo no debe usarse (por ejemplo, una lista de clases químicas o rangos de descriptores para los cuales se sabe que el modelo hace predicciones inexactas). Tales descripciones pueden proporcionarse por separado de la definición dada del AD, pero son de importancia similar. Muchas herramientas de software recientes proporcionan una evaluación automática de la aplicabilidad del modelo a la sustancia junto con el resultado de la predicción. En este caso, es importante que la evaluación automática pueda verificarse de manera independiente.

En el escenario preferible, la sustancia está dentro del dominio de aplicabilidad del modelo y las limitaciones específicas del modelo no se aplican a la sustancia.

Por el contrario, en algunos casos está claro que la sustancia está fuera del AD del modelo o que se aplican limitaciones específicas del modelo. En estos casos, la predicción no debe considerarse más (a menos que se proporcione una justificación válida, por ejemplo, no es técnicamente posible realizar una prueba experimental pero aún se necesita un valor numérico). Un ejemplo es un modelo cuyo dominio de aplicabilidad se define como un rango de propiedades fisicoquímicas, y además la descripción del modelo aclara que se ha desarrollado para predecir sustancias orgánicas dentro de ciertas clases químicas. En este caso, las predicciones de sustancias orgánicas fuera de estas clases no deben considerarse más, independientemente de si los valores fisicoquímicos están dentro del AD.

Entre estos dos escenarios claramente definidos hay casos intermedios. Un ejemplo de tales casos son sustancias para las cuales su relación con el dominio de aplicabilidad no se puede establecer de manera inequívoca, pero al mismo tiempo no cumplen con los criterios de exclusión por los cuales el modelo no debe usarse. Para estas sustancias, la precisión de la predicción del modelo puede ser desconocida, pero no necesariamente inaceptable para todos los fines regulatorios. Otro ejemplo son las predicciones para productos químicos que se encuentran cerca de los límites del dominio de aplicabilidad y, por lo tanto, pueden estar asociados con una mayor incertidumbre. Para estos casos, la evaluación de la confiabilidad de la predicción, como se describe en la siguiente sección, es decisiva para su aceptación general para un propósito regulatorio dado.

2.3 Fiabilidad de la(s) predicción(es)

Además del dominio de aplicabilidad (AD), se pueden considerar varios otros aspectos al evaluar la confiabilidad de una predicción. Si estos aspectos ya están incluidos en la definición del AD, la evaluación no necesita repetirse.

La Lista de Verificación de Predicción incluye los siguientes Elementos de Evaluación (EE) para verificar que una predicción sea confiable:

- Reproducibilidad
- Rendimiento general del modelo
- Ajuste dentro de los espacios fisicoquímicos, estructurales y de respuesta del conjunto de entrenamiento del modelo
- Rendimiento del modelo para sustancias similares
- Consideraciones mecanísticas y/o metabólicas
- Consistencia de la información

Reproducibilidad (EE 3.1 en las Listas de Verificación de Predicción y Resultados)

En primer lugar, se debe verificar que sea posible reproducir la predicción (es decir, obtener el mismo resultado) utilizando la entrada y el modelo documentados. Cuando los evaluadores tienen acceso al mismo modelo y versión utilizados para generar la predicción, pueden repetir la predicción para confirmar que se obtiene la misma predicción. Si se obtiene una predicción diferente sin una explicación válida (por ejemplo, una versión diferente del modelo), es indicativo de una menor confiabilidad.

Rendimiento general del modelo (EE 3.2 en las Listas de Verificación de Predicción y Resultados)

En segundo lugar, se debe considerar el rendimiento general del modelo. Representa la base para la evaluación de la confiabilidad y se espera que esté incluido en la documentación del modelo (QMRP). Como ejemplo, se puede esperar un error estándar similar para modelos que predicen resultados continuos (o la precisión de modelos que predicen resultados categóricos) para la predicción de la sustancia bajo análisis. Por esta razón, una predicción generada por un modelo con un mejor rendimiento debe considerarse más confiable que una predicción generada por un modelo con un rendimiento inferior.

Ajuste dentro de los espacios fisicoquímicos, estructurales y de respuesta del conjunto de entrenamiento del modelo (EE 3.3 en las Listas de Verificación de Predicción y Resultados)

Las sustancias en el conjunto de entrenamiento tienen valores definidos para los descriptores (como descriptores fisicoquímicos, descriptores moleculares, etc.) y para la propiedad de interés (valores de respuesta). Estos valores se pueden utilizar para definir espacios de descriptores y respuestas basados en rangos o distribuciones. Los evaluadores deben comparar los valores de descriptores y respuestas de la sustancia bajo análisis con los rangos definidos por el conjunto de entrenamiento.

Del mismo modo, las sustancias en el conjunto de entrenamiento se pueden utilizar para definir un espacio estructural en términos de grupos funcionales o fragmentos estructurales. Cuando todas las características estructurales de la sustancia son conocidas por el modelo, la predicción se considera más confiable. Las características estructurales que el modelo no conoce pueden tener un impacto en la confiabilidad de la predicción. Este aspecto es de particular importancia para las predicciones que indican la falta de actividad biológica en modelos basados en fragmentos o alertas, donde la falta de efecto puede deberse a la falta de conocimiento por parte del modelo, que podría no haber sido entrenado para predecir el efecto de ciertos fragmentos.

La confiabilidad de la predicción es menor cuando la sustancia bajo análisis cae en regiones de los espacios del conjunto de entrenamiento que están escasamente pobladas o asociadas con un menor rendimiento del modelo. Se debe prestar especial atención en tales casos.

Rendimiento del modelo para sustancias similares (EE 3.4 en las Listas de Verificación de Predicción y Resultados)

Una evaluación del rendimiento local del modelo requiere la identificación de sustancias similares con datos experimentales confiables para la propiedad de interés. Si no se realiza automáticamente por el software, este trabajo debe ser realizado por los usuarios de (Q)SAR y luego verificado por los evaluadores. Las sustancias similares pueden definirse considerando varios aspectos, como la similitud estructural, fisicoquímica y/o mecanística. Las sustancias similares pueden ser identificadas manualmente por los usuarios utilizando juicio experto, con o sin la ayuda de herramientas computacionales para la identificación de análogos como el "OECD QSAR Toolbox".

Algunos modelos (Q)SAR proporcionan automáticamente junto con la predicción una lista de sustancias similares a la entrada, sus valores experimentales y la precisión de sus predicciones.

Sin embargo, los índices de similitud molecular por sí solos no son necesariamente suficientes para definir la similitud, ya que algunos fragmentos moleculares pueden ser muy importantes para una propiedad, pero no para otra, y en algunos casos la co-ocurrencia de varios fragmentos puede ser muy importante (por ejemplo, agentes de entrecruzamiento).

Si hay disponibles datos experimentales confiables para sustancias similares a la que se está analizando, el rendimiento del modelo al predecir estas sustancias puede utilizarse para caracterizar mejor la confiabilidad de la predicción. El respaldo es mayor cuando las sustancias similares utilizadas para esta evaluación no forman parte del conjunto de entrenamiento del modelo, ya que este caso imita de manera más objetiva la aplicación en la vida real del modelo al predecir nuevas sustancias. Cuanto más similares sean las sustancias a la sustancia bajo análisis, especialmente en términos de descriptores, fragmentos u otras propiedades más relevantes para la predicción, más informativa será la precisión de sus predicciones para estimar la precisión para la predicción de la sustancia bajo análisis. En algunos casos, la mejora de la confiabilidad es aplicable incluso cuando la sustancia bajo análisis o sustancias similares están formalmente fuera del dominio de aplicabilidad del modelo. La sustancia bajo análisis puede estar fuera del dominio de aplicabilidad debido a la falta de datos experimentales en su espacio químico en el momento del desarrollo del modelo. Si nuevos datos para sustancias similares muestran un buen rendimiento del modelo en ese espacio químico, entonces hay una mejora concomitante en la confiabilidad de la predicción de la sustancia bajo análisis.

Para modelos basados en fragmentos/alertas, el rendimiento de las alertas positivas/negativas identificadas (expresado como, por ejemplo, estadísticas de Cooper (OECD, 2007)) también puede proporcionar información sobre el rendimiento local del modelo. En ausencia de datos para sustancias similares o cuando su identificación no es posible o factible, el resultado de este EE debe marcarse como "no aplicable/evaluado".

Consideraciones mecanísticas y/o metabólicas (EE 3.5 en las Listas de Verificación de Predicción y Resultados)

Este elemento de evaluación abarca consideraciones toxicocinéticas y toxicodinámicas relevantes para la propiedad bajo análisis. Una discusión sobre cómo las consideraciones sobre aspectos mecanísticos (como las Vías de Resultado Adverso, AOP) y las propiedades ADME3 (como las (bio)transformaciones, es decir, el metabolismo y otras transformaciones bióticas o abióticas como la hidrólisis, la autooxidación y la fotólisis) se ajustan a la predicción puede contribuir a la evaluación de la confiabilidad y, por lo tanto, debe proporcionarse cuando sea posible. Si se sabe o se predice que la estructura tiene un cierto mecanismo de acción relevante para la propiedad bajo análisis, o se espera la formación de ciertos metabolitos u otros productos de transformación, se espera que se describa su relación con la predicción y cómo el modelo los tiene en cuenta. Un ejemplo es la predicción de una propiedad para la cual se conocen modos de acción, como la sensibilización cutánea debido a la unión a proteínas. Si el modelo proporciona una predicción de "no sensibilizante", pero se sabe o se predice que la sustancia bajo análisis se une a proteínas, entonces se necesitan explicaciones adicionales para respaldar la confiabilidad de la predicción negativa. De manera similar, si se sabe que la sustancia bajo análisis sufre metabolismo que conduce a la formación de metabolitos peligrosos, una predicción de falta de peligro necesitaría una justificación adicional para discutir cómo la predicción del modelo considera los efectos del metabolismo y sus productos.

Consistencia de la información (EE 3.6 en las Listas de Verificación de Predicción y Resultados)

Finalmente, es común que las predicciones de más de un modelo se utilicen para predecir la propiedad de interés, y/o que haya información adicional (valores medidos o calculados) para la misma propiedad o una relacionada. La consistencia entre las predicciones de (Q)SAR y/o con otra información confiable, si está disponible, debe ser considerada. Las predicciones consistentes (es decir, en acuerdo) entre sí se consideran más confiables cuando son generadas por modelos independientes (en términos de conjuntos de entrenamiento, técnicas de modelado y/o descriptores/alertas utilizados). La información contradictoria, en ausencia de alguna explicación, tiende a disminuir la confiabilidad de la predicción. La Lista de Verificación de Resultados debe utilizarse cuando un resultado de (Q)SAR se basa en múltiples predicciones (ver Capítulo 3). Consideraciones adicionales de peso de la evidencia están fuera del alcance de este documento.

La evaluación de la confiabilidad se anticipa como la parte más compleja de la evaluación. Requiere un análisis en profundidad y juicio experto, incluso cuando se facilite con los informes exhaustivos de los programas modernos de (Q)SAR. Sin embargo, cuando se realiza una evaluación de confiabilidad específica para una predicción de manera exhaustiva, puede ofrecer más información que la información más genérica sobre el dominio de aplicabilidad del modelo.

2.4 Adecuación del resultado para el propósito regulatorio

Este principio describe la evaluación de la utilidad de la predicción de (Q)SAR para responder a una pregunta regulatoria específica. Incluso si la predicción o el resultado de múltiples predicciones son muy confiables, solo pueden usarse para un propósito regulatorio específico si son adecuados para ese propósito. Los elementos a considerar al evaluar la idoneidad para el propósito dependen del marco regulatorio. La decisión final sobre la idoneidad para el propósito (y sobre la aceptabilidad en general) recae en la autoridad responsable del marco regulatorio.

La Lista de Verificación de Predicciones incluye los siguientes Elementos de Evaluación (EE) para verificar si un resultado es adecuado para el propósito regulatorio:

- Cumplimiento con requisitos adicionales.
- Correspondencia entre la propiedad predicha y la propiedad requerida por la regulación.
- Decidibilidad dentro del marco específico.

Estos elementos buscan asegurar que la predicción o el resultado de múltiples predicciones sean coherentes con los requisitos y objetivos específicos del marco regulatorio en el que se están utilizando. La evaluación de la idoneidad para el propósito considera si la información proporcionada por la predicción es relevante, confiable y adecuada para tomar decisiones en el contexto regulatorio en cuestión.

Cumplimiento con requisitos adicionales (EE 4.1 en las Listas de Verificación de Predicciones y Resultados)

Un marco regulatorio puede indicar directamente criterios específicos para el uso aceptable de las predicciones de (Q)SAR. Si los criterios incluyen elementos que no están cubiertos en este documento, se espera que la predicción cumpla con estos criterios adicionales para considerar su uso como aceptable. Como ejemplo, en el caso de regulaciones que abordan sustancias con composiciones complejas o mezclas, un elemento crucial para evaluar es la consideración de toda la composición de la sustancia o componentes de la mezcla siguiendo los criterios especificados por la regulación (por ejemplo, teniendo en cuenta todos los componentes presentes a una concentración por encima de un umbral dado, utilizando un enfoque de adición de concentración o dosis, o seleccionando un caso peor razonable).

En algunos casos, también puede ser necesario considerar los efectos antagónicos y sinérgicos potenciales de los diferentes componentes. Otro ejemplo son las regulaciones que requieren el uso combinado de dos tipos diferentes de modelos para producir un resultado (como la guía ICH M7 para la evaluación y control de impurezas reactivas al ADN (mutagénicas) en productos farmacéuticos para limitar el riesgo potencial de cáncer, que requiere el uso de un modelo basado en reglas de expertos y otro basado en estadísticas). En este caso, el uso de una predicción individual no será adecuado para el propósito.

Correspondencia entre la propiedad predicha y la propiedad requerida por la regulación (EE 4.2 en las Listas de Verificación de Predicciones y Resultados)

Es importante que la propiedad predicha por el modelo de (Q)SAR coincida con la propiedad requerida por la regulación. La propiedad predicha por un modelo de (Q)SAR depende de los datos experimentales aplicados como conjunto de entrenamiento y de la curación de datos realizada en los datos antes del modelado de (Q)SAR (por ejemplo, criterios de inclusión/exclusión, transparencia de evaluaciones manuales y eliminación de valores atípicos, umbrales aplicados para definir positivos y negativos, etc.). Como ejemplo, cepas bacterianas específicas y la presencia de activación metabólica pueden necesitar ser considerados explícitamente por el modelo si lo requiere la regulación al evaluar la mutagenicidad in vitro en un ensayo de mutación reversa bacteriana. Si la regulación se refiere a una guía de prueba específica, el modelo debe incluir los resultados experimentales obtenidos siguiendo la guía de prueba especificada en su conjunto de entrenamiento. Sin embargo, esto no siempre puede ser necesario, por ejemplo, los modelos para la mutagenicidad Ames pueden incluir datos históricos que no se realizaron utilizando todas las cepas actualmente requeridas, pero las predicciones positivas del modelo aún pueden ser adecuadas. Otro ejemplo es la vida media de degradación, donde las predicciones de modelos entrenados con datos de vida media de degradación primaria no son adecuadas para evaluar la vida media de degradación última.

Decidibilidad dentro del marco específico (EE 4.3 en las Listas de Verificación de Predicciones y Resultados)

Una predicción es adecuada para su propósito cuando permite tomar una decisión regulatoria en el marco de su uso. En presencia de un umbral regulatorio, la predicción será "decidible" cuando sea comparable al umbral (es decir, expresado en la misma unidad o en una unidad convertible) y proporcione un nivel suficiente de confianza de que la predicción se encuentra en un lado del umbral o lo suficientemente cerca como para no descartar una preocupación. Como ejemplo, si un modelo predice una solubilidad en agua de 0.09 mg/L con una alta incertidumbre y la decisión regulatoria depende de si la sustancia tiene una solubilidad por encima o por debajo de 0.10 mg/L, entonces la predicción no será "decidible". El nivel de confianza requerido, así como la forma de establecerlo, dependerá del propósito regulatorio previsto.

2.5. Conclusión sobre la evaluación de una predicción individual.

La conclusión de la Lista de Verificación de Predicciones consta de un valor de incertidumbre para la evaluación y un resultado basado en esta incertidumbre.

2.5.1. Incertidumbre

Cuando se utiliza la Lista de Verificación de Predicciones, los evaluadores deben decidir si cada EE se cumple o no. En la mayoría de los casos, esta decisión incluirá un cierto nivel de incertidumbre.

La incertidumbre de cada EE aplicable puede describirse utilizando valores semicuantitativos (bajo, medio o alto), mientras que los EEs que no son aplicables/evaluados no reciben un puntaje de incertidumbre. Para los EEs que se cumplen, la explicación de cómo asignar el nivel de incertidumbre para cada EE se proporciona en una tabla separada de la Lista de Verificación de Predicciones. Además, se asigna automáticamente una alta incertidumbre predeterminada a los EEs que no se cumplen o no se documentan, a menos que se proporcione una justificación válida. Luego, la incertidumbre asociada con cada EE se puede utilizar para decidir sobre la incertidumbre general de la predicción.

La incertidumbre general de la Lista de Verificación de Predicciones se considera "Baja" cuando:

- la predicción cumple con todos los elementos con baja incertidumbre, O
- la mayoría de los elementos tienen baja incertidumbre y los elementos con mayor incertidumbre (incluidos los elementos no documentados o no cumplidos, si los hay) no se consideran de alto peso en la evaluación general, y se proporciona una justificación convincente de esta consideración.

La incertidumbre general de la Lista de Verificación de Predicciones se considera "Media" cuando:

- la predicción cumple con todos o la mayoría de los elementos con incertidumbre media. Los elementos con mayor incertidumbre (si los hay) no se consideran de alto peso en la evaluación general, o
- la predicción cumple con la mayoría de los elementos con baja incertidumbre, pero algunos elementos de alto peso en la evaluación general tienen incertidumbre media.

La incertidumbre general de la Lista de Verificación de Predicciones se considera "Alta" cuando:

- la predicción cumple con todos o la mayoría de los elementos con alta incertidumbre, o
- la predicción cumple con la mayoría de los elementos con baja o media incertidumbre, pero algunos elementos de alto peso en la evaluación general tienen alta incertidumbre.

2.5.2. Resultado

El resultado de la evaluación de la predicción se basa en la integración del resultado y la incertidumbre de los EEs en la Lista de Verificación de Predicciones. En general, los EEs con mayor peso son los más críticos para la evaluación, y una predicción aceptable debería cumplir todos ellos con una incertidumbre baja o media. Los EEs con un peso menor también son importantes, pero las predicciones que cumplen algunos de ellos con alta incertidumbre o que no los cumplen también pueden ser aceptables para algunas aplicaciones. Siguiendo la lógica descrita en el Capítulo 2.5.1., este enfoque corresponde a considerar que las predicciones con incertidumbre Baja o Media son aceptables.

Este documento proporciona consejos generales sobre la evaluación y sugiere el "peso" que se puede utilizar para cada EE, pero cada autoridad podría establecer (y comunicar) requisitos diferentes para resultados aceptables según sus aplicaciones.

Para una predicción individual, la evaluación se concluye en esta etapa.

3.1 Cuándo utilizar la Lista de Verificación de Resultados

La Lista de Verificación de Resultados debe utilizarse al evaluar un resultado derivado de múltiples predicciones para propiedades iguales o relacionadas. Los casos que involucran múltiples predicciones incluyen:

- a. Predicciones de diferentes modelos para la misma estructura;
- b. Predicciones del mismo modelo para diferentes estructuras (como los múltiples constituyentes de una sustancia o para la sustancia en análisis y sus metabolitos o productos de transformación);
- c. Una combinación de los casos anteriores.

Primero, cada predicción debe evaluarse utilizando las listas de verificación de predicciones individuales dentro de la Lista de Verificación de Resultados. Para casos complejos (punto c), se recomienda comenzar abordando las múltiples predicciones asociadas con la misma estructura, y luego considerar las predicciones para diferentes estructuras.

Se debe completar una lista de verificación para cada predicción (dentro de la Lista de Verificación de Resultados). La evaluación de cada predicción puede realizarse de manera independiente, excepto para los Elementos de Evaluación (EEs) enumerados a continuación. Para estos EEs, el resultado debe considerar la información de todas las predicciones:

- Entrada representativa de la sustancia en análisis;
- Consideraciones mecánísticas y/o metabólicas;
- Consistencia de la información;
- Todos los EEs que hacen referencia al principio "El resultado es adecuado para el propósito regulatorio".

Esta situación se puede ejemplificar con un caso en el que se predice individualmente dos componentes de la misma sustancia. Deben compilarse dos listas de verificación dentro de la Lista de Verificación de Resultados, una para cada predicción. Al evaluar el EE sobre la representatividad de la entrada, se deben considerar ambos componentes, y el mismo resultado puede registrarse en las listas de verificación para cada predicción dentro de la Lista de Verificación de Resultados.

Luego, hay un EE adicional que se debe considerar al evaluar un resultado de múltiples predicciones:

- Correcta determinación del resultado final a partir de predicciones individuales

Los evaluadores deben verificar que el resultado final, derivado de los valores individuales predichos, se haya determinado correctamente. Dependiendo del tipo de punto final y el requisito regulatorio, el resultado final puede determinarse por consenso mayoritario, caso peor, valor promedio o técnicas más complejas. En este cálculo, pueden asignarse diferentes pesos a las predicciones según su confiabilidad, por ejemplo. En cualquier caso, los evaluadores deben verificar que la determinación esté documentada y adecuadamente justificada.

3.2 Incertidumbre y resultado de la predicción (Q)SAR

3.2.1. Incertidumbre

Finalmente, hay un paso adicional destinado a ponderar cada predicción con el fin de llegar a una conclusión, es decir, decidir sobre la aceptabilidad del resultado final. La lógica para la estimación de la incertidumbre del resultado derivado de múltiples predicciones es la siguiente:

- Cuando el resultado final se genera mediante la integración de predicciones consistentes (es decir, en acuerdo) para la misma estructura provenientes de diferentes modelos independientes, entonces la incertidumbre del resultado final puede ser igual o menor que la incertidumbre de las predicciones individuales. La incertidumbre de un resultado de múltiples predicciones puede ser menor cuando los EE con mayor incertidumbre son diferentes para las predicciones y, por lo tanto, pueden respaldarse mutuamente en sus elementos de incertidumbre. Esto no ocurre cuando una mayor incertidumbre proviene de los mismos EE;
- Cuando el resultado final se genera mediante la integración de predicciones inconsistentes para la misma estructura provenientes de modelos independientes, entonces la incertidumbre del resultado final será igual o mayor que la incertidumbre de las predicciones individuales. Normalmente, los resultados derivados de predicciones inconsistentes

serán aceptables solo si una o más predicciones en acuerdo superan las inconsistencias debido a un mejor ajuste al dominio de aplicabilidad, mayor confiabilidad o menor incertidumbre. Se espera una justificación sobre la base de la integración por parte del usuario de (Q)SAR, mientras que el evaluador de (Q)SAR puede comentar sobre este aspecto en el EE "Consistencia de la información";

- Cuando el resultado final depende de predicciones para diferentes estructuras (por ejemplo, diferentes constituyentes de la misma sustancia y / o un compuesto padre y sus metabolitos, etc.), entonces su incertidumbre debe tener en cuenta la incertidumbre de las predicciones individuales y la composición de la sustancia bajo análisis. Como ejemplo, al predecir la solubilidad en agua, la incertidumbre de la predicción para el constituyente principal podría tener una influencia mayor en la incertidumbre del resultado final en comparación con la incertidumbre de la predicción para una impureza menor. En este ejemplo, si la incertidumbre de la predicción para el constituyente principal es "Media" y la incertidumbre para la predicción de la impureza menor es "Alta", entonces la incertidumbre del resultado final podría ser "Media". Mientras que, si la incertidumbre de la predicción para el constituyente principal es "Alta", mientras que para la impureza menor es "Media", entonces la incertidumbre del resultado final podría ser "Alta". Para otras propiedades, como la mutagenicidad, donde una impureza mutagénica puede hacer que toda la sustancia sea mutagénica, la concentración de un constituyente o impureza en la composición tendrá un papel más pequeño o ningún papel en ponderar las predicciones individuales para decidir sobre la incertidumbre del resultado general.

3.2.2. Resultado

La incertidumbre del resultado puede utilizarse para determinar el resultado de la evaluación. Coherente con el enfoque para las predicciones individuales descrito en el numeral 2.5.2, los resultados con una incertidumbre Baja o Media podrían considerarse aceptables. La decisión del resultado para el resultado final completa la evaluación.

4 Consideraciones Finales

Este documento tiene como objetivo proporcionar una lista exhaustiva de los elementos necesarios para evaluar modelos (Q)SAR, predicciones y resultados de múltiples predicciones. Dependiendo de las especificidades del caso y el contexto de uso, es posible que no todos los EE sean aplicables o requeridos. Cada autoridad reguladora debe decidir y comunicar qué elementos deben considerarse sistemáticamente para la evaluación en un marco regulatorio y un contexto específicos (por ejemplo, los requisitos de confiabilidad pueden ser más bajos cuando los (Q)SAR se utilizan con el propósito de cribar una base de datos en comparación con cuando se utilizan para evaluar una sustancia individual). Es posible que algunas autoridades introduzcan umbrales para juzgar si se cumple un EE, pero no se incluyen en este documento por al menos dos razones: 1. falta de umbrales generalmente acordados en la comunidad científica y 2. los umbrales aceptables pueden variar según una multitud de factores, incluida la propiedad de interés y el propósito o contexto regulatorio de uso.

En general, se debe esperar el uso de modelos (Q)SAR válidos. Esto se puede verificar mediante la Lista de Verificación del Modelo.

Siempre se debe requerir una entrada correcta, independientemente del contexto de uso. Si la entrada no es correcta, los (Q)SAR no serán útiles para consideraciones regulatorias adicionales.

La relación de la sustancia con el AD del modelo también es importante. Idealmente, la sustancia debería estar claramente dentro del AD del modelo. Si no es así, otros aspectos de confiabilidad aún pueden llevar a la aceptación del uso de la predicción o el resultado para un propósito determinado.

Finalmente, incluso si son correctas y confiables, una predicción o resultado solo se puede utilizar si es adecuada para el propósito regulatorio. Los requisitos regulatorios claros para el uso de resultados (Q)SAR facilitan la evaluación de este principio.

Cuando se integran múltiples predicciones en un resultado, es esencial proporcionar una descripción y una justificación para el enfoque utilizado para obtener el resultado general.

Este documento incluye una actualización importante del formato QPRF. También se ha identificado la necesidad de una actualización más completa del formato QMRF, que se espera que comience después de la publicación de este documento. Además, se está considerando el desarrollo de una nueva plantilla para informar resultados basados en múltiples predicciones.

Las Listas de Verificación del QAF ofrecen explicaciones y ejemplos adicionales útiles para realizar la evaluación en la práctica. Después de la publicación de este documento, es posible que se preparen guías o estudios de casos específicos para regulaciones para aclarar los requisitos específicos de predicción y regulación.