

Validación, Calificación y Aceptación Regulatoria de Metodologías de Nuevo Enfoque

Marzo, 2024

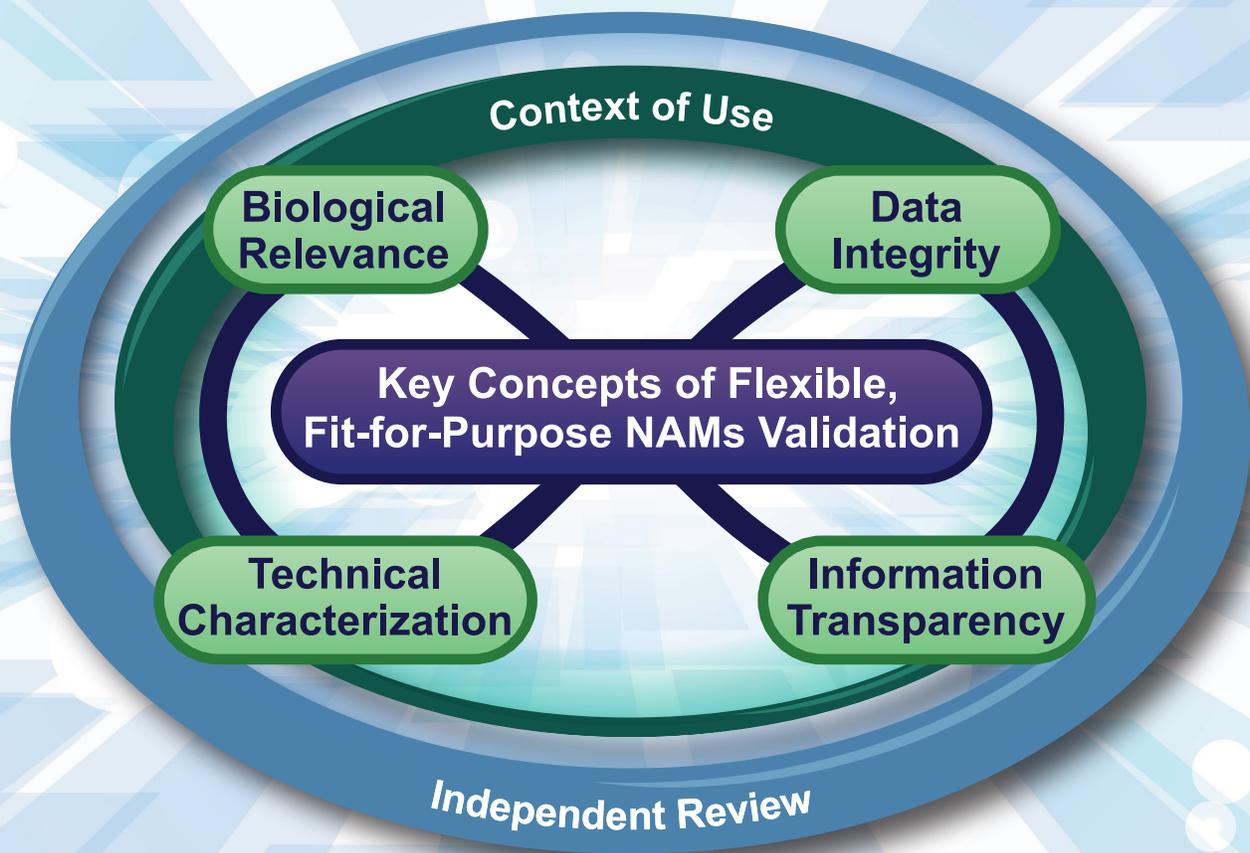


Tabla de contenido

Lista de Tablas	v
Lista de figuras.....	vi
Abreviaturas y Acrónimos	vii
Grupo de Trabajo de Validación del ICCVAM.....	ix
Representantes de la Agencia ICCVAM	xi
Descargo de responsabilidad.....	1
Resumen Ejecutivo	1
1.0 Introducción	2
2.0 Conceptos clave en la validación de MNE flexibles y aptas para el propósito pretendido	4
3.0 Aplicación de los conceptos clave para generar confianza en las MNE.....	6
3.1 Contexto de Uso	6
3.2 Relevancia biológica	8
3.2.1 Comprensión mecanicista	8
3.2.2 Compuestos de referencia	11
3.2.3 Comparación con Métodos Existentes en Animales de Laboratorio	14
3.3 Caracterización Técnica	17
3.3.1 Incorporación de Herramientas de Calidad Seleccionadas	17
3.3.2 Mejores Prácticas de Control de Calidad.....	20
3.3.3 Documentación	25
3.4 Integridad de los datos.....	29
3.5 Transparencia de la Información.....	29
3.6 Revisión Independiente	30
4.0 Aceptación de MNE por Agencias Federales de los Estados Unidos.....	31

4.1	Comprensión de las Necesidades Regulatorias y Contextos de Decisión.....	31
4.2	Consideraciones sobre el Contexto de Uso	32
4.3	Evolución de la Confianza Basada en la Experiencia Adquirida.....	32
5.0	Estados Unidos y la armonización internacional	33
5.1	Armonización de Estados Unidos: papel del ICCVAM y del NICEATM	33
5.2	Armonización de Estados Unidos: Colaboraciones Federales Adicionales para Avanzar las 3Rs.....	34
5.3	Armonización Internacional.....	35
6.0	Comunicación y Capacitación para Fomentar el Uso de MNE	36
7.0	Conclusión e implementación.....	37
	Referencias.....	38
	APÉNDICE A: GLOSARIO.....	49
	APÉNDICE B: HERRAMIENTAS DE CALIDAD	55
1.0	Diagramas de flujo	55
2.0	Análisis de Causa-Efecto	56
3.0	Gráficas de control	57
4.0	Hojas de Verificación.....	59
5.0	Gráficas de Dispersión	59
	APÉNDICE C: EVALUACIÓN DE MÉTODOS ANALÍTICOS	61
1.0	Límites de Detección y Cuantificación	61
2.0	Identificación de Interferencias	61
3.0	Evaluación de la Precisión Analítica.....	61
4.0	Estabilidad de los Materiales.....	61
5.0	Pruebas de Robustez.....	62

6.0	Análisis de Recuperación	63
7.0	Análisis Técnico del Dominio de Aplicabilidad	63
8.0	Control Positivo.....	64
9.0	Estándares de Referencia para Calibración de Instrumentos	64
10.0	Configuración de Especificaciones	65

Lista de Tablas

Tabla 1. Ejemplos de documentos de Estados Unidos e internacionales existentes, relacionados con la validación, calificación y uso regulatorio de MNE.....	3
Tabla 2. Manuscritos producidos por grupos de trabajo del ICCVAM, que proporcionan detalles sobre las necesidades de pruebas de la agencia para temas seleccionados.....	7
Tabla 3. Ejemplos de criterios de valoración en los que se ha demostrado la relevancia biológica y mecanicista de las MNE	9
Tabla 4. Usos Seleccionados de Compuestos de Referencia.....	12
Tabla 5. Herramientas de Calidad Utilizadas con Frecuencia en la Caracterización Técnica de las MNE.....	19
Tabla 6. Evaluación de Métodos Analíticos	22

Lista de figuras

Figura 1. Conceptos clave a considerar durante el desarrollo e implementación de estrategias de validación de MNE flexibles y aptas para su propósito.....	5
Figura 2. Entorno para el desarrollo de MNE robustas.	18
Figura suplementaria 1. Diagrama de flujo que describe el protocolo del MTS modificado....	55
Figura Suplementaria 2. Diagrama de causa-efecto	56
Figura suplementaria 3. Datos de la gráfica de control	58
Figura complementaria 4. Correlación de los valores de CE ₅₀ del sulfato de cadmio.	60

Abreviaturas y Acrónimos

Las 3Rs: reemplazo, reducción o refinamiento del uso de animales

RRA: ruta de resultados adversos

C&E: causa-efecto

CDU: contexto de uso

CPSC: Comisión de Seguridad de Productos para el Consumidor de los Estados Unidos (según sus siglas en inglés)

ED: enfoque definido

EPA: Agencia de Protección Ambiental de los Estados Unidos (según sus siglas en inglés)

EURL ECVAM: Laboratorio de Referencia de la Unión Europea para la Validación de Métodos Alternativos (según sus siglas en inglés)

FDA: Administración de Alimentos y Medicamentos de los Estados Unidos, según sus siglas en inglés

DG: Documento de Guía

GHS: Sistema Globalmente Armonizado de Clasificación y Etiquetado de Sustancias Químicas, según sus siglas en inglés

GIVIMP: Documento de Guía sobre las Buenas Prácticas de Métodos In Vitro, según sus siglas en inglés

BPL: Buenas Prácticas de Laboratorio

IATA: Enfoques Integrados para Pruebas y Evaluaciones, según sus siglas en inglés

ICATM: Cooperación Internacional sobre Métodos de Prueba Alternativos, según sus siglas en inglés

ICCVAM: Comité de Coordinación Interinstitucional para la Validación de Métodos Alternativos, según sus siglas en inglés

ICH: Consejo Internacional para la Armonización de los Requisitos Técnicos para Productos Farmacéuticos de Consumo Humano, según sus siglas en inglés

IQ/OQ/PQ: Calidad de Instalación/Calidad de Operación/Calidad de Desempeño, según sus siglas en inglés

JaCVAM: Centro Japonés para la Validación de Métodos Alternativos, según sus siglas en inglés

MTS: (3-(4,5-dimetiltiazol-2-il)-5-(3-carboximetoxifenil)-2-(4-sulfofenil)-2H-tetrazolio)

MNE: Metodologías de Nuevo Enfoque, según sus siglas en inglés

NCATS: Centro Nacional para el Avance de las Ciencias Traslacionales, según sus siglas en inglés

NICEATM: Centro Interinstitucional del Programa Nacional de Toxicología para la Evaluación de Métodos Toxicológicos Alternativos, según sus siglas en inglés

NIEHS: Instituto Nacional de Ciencias de la Salud Ambiental, según sus siglas en inglés

DO: densidad óptica

OCDE: Organización para la Cooperación y el Desarrollo Económicos

PNO: procedimiento normalizado de operación

GP: guía de prueba

Tox21: Toxicología en el siglo XXI

TSAR: Sistema de Seguimiento de Métodos Alternativos con miras a la Aceptación Regulatoria, según sus siglas en inglés

Grupo de Trabajo de Validación del ICCVAM

Agencia de Registro de Enfermedades y Sustancias Tóxicas

Moiz Mumtaz, PhD

Comisión de Seguridad de Productos para el Consumidor

John Gordon, PhD (Copresidente)

Departamento de Defensa

Donald Cronce, PhD

Natalia García-Reyero Vinas, Doctora

Matthew Johnson, DVM, DACLAM (a junio del 2022)

Emily N. Reinke, PhD (a abril del 2022)

Departamento de Asuntos de Veteranos

George Lathrop, Jr., DVM, MS, DACLAM

Agencia de Protección Ambiental

Oficina de Programas de Pesticidas

Anna Lowit, PhD

Scott Lynn, PhD

Monique Perron, ScD

Oficina de Investigación y Desarrollo

Kelly Carstens, PhD

Alison Harrill, PhD

Nisha Sipes, PhD

Administración de Alimentos y Medicamentos

Centro de Dispositivos y Salud Radiológica

Jennifer Goode, BS

Centro de Evaluación e Investigación de Medicamentos

Paul C. Brown, PhD

Centro de Seguridad Alimentaria y Nutrición Aplicada

Suzanne Fitzpatrick, PhD (Copresidente)

Anneliese Striz, PhD

Centro de Productos de Tabaco

Jueichuan (Connie) Kang, PhD

Oficina del Científico en Jefe

Tracy Chen, PhD

Instituto Nacional de Ciencias de la Salud Ambiental

Warren Casey, PhD

Centro Interinstitucional del Programa Nacional de Toxicología para la Evaluación de Métodos Toxicológicos Alternativos (NICEATM, según sus siglas en inglés)

Helena Hogberg, PhD

Nicole Kleinstreuer, PhD

Institutos Nacionales de Salud

Centro Nacional para el Avance de las Ciencias Traslacionales

Menghang Xia, PhD

Instituto Nacional de Estándares y Tecnología

Elijah Petersen, PhD (Copresidente)

Administración de Salud y Seguridad Laboral

Janet Carter (MS)

Personal de Soporte Contratado del NICEATM (Inotiv)

David Allen, PhD (a diciembre del 2023)

Michaela Blaylock

Amber Daniel, MTOx

Agnes Karmaus, PhD (a abril del 2023)

Representantes de la Agencia ICCVAM

Miembro del comité ICCVAM en el momento de publicación del presente informe

*Representante principal de la agencia; +Representante principal alternativo de la agencia

Agencia de Registro de Enfermedades y Sustancias Tóxicas

- * Moiz Mumtaz, PhD
- Patricia Ruiz, PhD

Comisión de Seguridad de Productos para el Consumidor de los Estados Unidos

- * John Gordon, PhD
- + Kristina Hatlelid, PhD, MPH
- + Eric Hooker, MS
- + Joanna Matheson, PhD

Ministerio de Agricultura de los Estados Unidos

- * Jessie Carder
- + Patrice Klein, MS, VMD, DACPV, DACVPM
- Erika Edwards
- Ben Green, PhD
- Katherine Horak, PhD

Ministerio de Defensa de los Estados Unidos

- * Shannon Marko, DVM, DACLAM
- + Natalia Garcia-Reyero Vinas, PhD (Copresidente)
- + Saber Hussain, PhD, Becario de ATS, Becario de AFRL
- + Elaine Merrill, PhD
- Matthew Grogg, PhD
- Donald Cronce, PhD

Ministerio de Energía de los Estados Unidos

R. Todd Anderson, PhD

Ministerio del Interior de los Estados Unidos

- * Barnett A. Rattner, PhD
- + Jessica K. Leet, PhD

Ministerio de Transporte de los Estados Unidos

- * Steve Hwang, PhD
- + Rebecca Rothhaas
- + Ryan Vierling, PhD

Oficina de Investigación y Desarrollo del Ministerio de Asuntos de Veteranos de los Estados Unidos

- * Holly Krull, PhD
- + George Lathrop, Jr., DVM, MS, DAACLAM

Agencia de Protección Ambiental de los Estados Unidos

Oficina de Seguridad Química y Prevención de Contaminación Ambiental

Charles Kovatch

Oficina de Programas de Pesticidas

Monique Perron, PhD

William (Bill) Eckel, PhD

Cecilia Tan, PhD

Oficina de Prevención de Contaminación Ambiental y Sustancias Tóxicas

- * Anna Lowit, PhD
- Louis (Gino) Scarano, PhD

Oficina de Investigación y Desarrollo

- + Alison Harrill, PhD
- Kelly Carstens, PhD
- Grace Patlewicz, PhD

Administración de Alimentos y Medicamentos

Centro de Evaluación e Investigación de Productos Biológicos

Leslie Wagner

Allen Wensky, PhD

Centro de Dispositivos y Salud Radiológica

- + Jennifer Goode, BS
- Simona Bancos, PhD
- Rakhi M. Dalal-Panguluri, PhD

Centro de Evaluación e Investigación de Medicamentos

Paul C. Brown, PhD

Nakissa Sadrieh, PhD

Centro de Seguridad Alimentaria y Nutrición Aplicada

- * Suzanne Fitzpatrick, PhD (Copresidente)
- Omari J. Bandele, PhD
- Patrick Crittenden, PhD
- Shruti Kabadi, PhD

Centro de Productos de Tabaco

Jueichuan (Connie) Kang, PhD

Centro de Medicina Veterinaria

M. Cecilia Águila, DVM
Li You, PhD

Centro Nacional de Investigaciones Toxicológicas

Mugimane (Manju) Manjanatha, PhD

Oficina del Científico en Jefe

Tracy Chen, PhD
Chad P. Nelson, PhD, MSPH

Instituto Nacional del Cáncer

- * Brian Cholewa, PhD
- + Ron Johnson, PhD
- Gary Robinson, PhD

Instituto Nacional de Salud y Seguridad Laboral

- * Stephen Leonard, PhD

Instituto Nacional de Ciencias de la Salud Ambiental

- * Warren Casey, PhD
- + Stephen Ferguson, PhD
- + David M. Reif, PhD
- Nicole Kleinstreuer, PhD

Instituto Nacional de Estándares y Tecnología

- * John Elliott, PhD
- + Elijah Petersen, PhD

Institutos Nacionales de Salud

- * Nicolette Petervary, VMD, MS, DACAW
- Michael Eichner, DVM, DACLAM

Biblioteca Nacional de Medicina

- * Dina N. Paltoo, PhD, MPH, CPI

Administración de Salud y Seguridad Laboral

- * Deana Holmes, MT
- + Janet Carter, MS

Descargo de responsabilidad

Este informe ha sido desarrollado como un recurso para agencias federales estadounidenses y partes interesadas que buscan establecer confianza en nuevos enfoques que reemplacen, reduzcan o refinen el uso de animales en pruebas. Los principios descritos en este informe se desarrollaron con insumos de personal de 17 dependencias federales, múltiples grupos de trabajo interinstitucionales, el público y el Comité Científico Asesor sobre Métodos Toxicológicos Alternativos. Como tal, este informe no refleja necesariamente las opiniones o políticas de ninguna agencia o grupo de trabajo. Este documento no crea derechos para ninguna persona o parte y no debe ser considerado como compromiso por ningún organismo federal. El presente informe no establece ninguna responsabilidad exigible por ley. Más bien, este informe describe la postura actual del Comité Coordinador Interinstitucional para la Validación de Métodos Alternativos (ICCVAM) sobre un tema y se deberá considerar solo como recomendaciones, a menos que se citen requisitos regulatorios o legales específicos. El uso de la palabra “deberá” significa que se sugiere o recomienda algo, pero no constituye un requisito.

Resumen Ejecutivo

Cada vez más se están desarrollando Metodologías de Nuevo Enfoque (MNE) y se están utilizando para proporcionar evaluaciones reglamentarias y no reglamentarias de los posibles efectos tóxicos de sustancias y productos químicos en la salud humana y el medio ambiente. Estas MNE se están utilizando para investigar los mecanismos biológicos subyacentes en procesos toxicológicos, para ayudar en la evaluación de productos nuevos y existentes, y para generar información de identificación de peligros y relación dosis-respuesta, para los propósitos de clasificación de peligros para la salud y el medio ambiente y evaluación de riesgos. El presente informe fue desarrollado por el ICCVAM para ayudar a los desarrolladores y usuarios finales a generar confianza en las MNE. Esta confianza se puede lograr mediante la implementación de estrategias de validación flexibles y adecuadas, que tomen en cuenta la aplicación prevista de la MNE. Este documento ayuda a generar confianza al describir conceptos como el contexto de uso, la importancia biológica y la caracterización técnica de las MNE. El informe busca la aceptación regulatoria de las MNE por parte de las agencias federales de Estados Unidos y la potencial armonización internacional de las mismas, y revisa las mejores prácticas para las herramientas de calidad y la evaluación técnica de las MNE. El informe también enfatiza la necesidad de comunicación entre los desarrolladores, los usuarios finales y las agencias reguladoras. La estrategia obsoleta de la validación de una sirve para todas, no funciona en un campo avanzado como las MNE. El ICCVAM es una organización que puede apoyar esta transición hacia enfoques más modernos, en calidad de asesor, a través de la coordinación de los esfuerzos de validación y estableciendo oportunidades de colaboración en investigación. Por lo tanto, el ICCVAM ha desarrollado este informe, en cooperación con 17 agencias reguladoras federales y laboratorios de investigación, para ayudar a los desarrolladores, usuarios finales y agencias reguladoras a generar confianza en las MNE, para que puedan implementarse de modo que complementen o reemplacen a las pruebas en animales con fines regulatorios y no regulatorios.

1.0 Introducción

El Comité Coordinador Interinstitucional para la Validación de Métodos Alternativos (ICCVAM, según sus siglas en inglés) está integrado por representantes de 17 organismos reguladores federales y de investigación de los Estados Unidos, que requieren, utilizan, generan o difunden, información sobre pruebas toxicológicas y de seguridad. El ICCVAM realiza evaluaciones técnicas de métodos de prueba de seguridad nuevos, revisados y alternativos, y estrategias de prueba integradas con aplicabilidad regulatoria. El ICCVAM también promueve la validación científica y la aceptación o calificación regulatoria, de métodos de prueba que evalúan con precisión la seguridad química y los peligros de productos relevantes, en un esfuerzo por reemplazar, reducir o refinar (mejorar el bienestar de los animales y disminuir o evitarles dolor y angustia) el uso de animales.

Poco después de su establecimiento como comité permanente en 1997, el ICCVAM publicó un informe, “Validación y Aceptación Regulatoria de Métodos de Pruebas Toxicológicas” (ICCVAM, 1997), en el que se señalaban criterios para la validación y aceptación regulatoria de métodos de prueba nuevos y alternativos. Posteriormente se proporcionó orientación adicional en la publicación de 2003, “Directrices del ICCVAM para la Nominación y Presentación de Métodos de Prueba Nuevos, Revisados y Alternativos” (ICCVAM, 2003). Los principios señalados en estos documentos se llevaron más allá en la elaboración de los criterios internacionales descritos en el Documento de Orientación (GD) 34 de la Organización para la Cooperación y el Desarrollo Económicos (OCDE), “Documento de Orientación sobre la Validación y Aceptación Internacional de Métodos de Prueba Nuevos o Actualizados para la Evaluación de Peligros” (OCDE, 2005). Estos recursos describían un modelo de validación que es flexible en principio, pero que en la práctica ha mostrado varias limitaciones, como ser largo y consumir muchos recursos. Para algunos contextos de uso, es posible que los métodos no necesiten someterse a cada uno de los pasos de este proceso de validación, para que produzcan datos valiosos para una agencia federal. Además, estos documentos no siempre son compatibles con muchos enfoques modernos de pruebas de toxicidad, que ponen menos énfasis en la sustitución de las pruebas *in vivo* por un único método alternativo y más énfasis en la integración de resultados provenientes de múltiples ensayos *in vitro e in químico* y enfoques *in silico* (por ejemplo, modelos computacionales). La publicación del ICCVAM del 2018, “Hoja de ruta estratégica para establecer nuevos enfoques para evaluar la seguridad de sustancias químicas y productos médicos en los Estados Unidos” (ICCVAM, 2018), proporciona un marco conceptual que promueve una mejor comunicación entre agencias y desarrolladores de métodos de prueba y una mayor flexibilidad en el modo en que se establece la confianza, para ayudar a garantizar la adopción de nuevos métodos por parte de las agencias federales y las industrias reguladas, una vez validados para una aplicación o contexto de uso (CDU) específico. El texto que sigue proporciona información más específica sobre el establecimiento de confianza en las metodologías de nuevo enfoque (MNE) basadas en los principios descritos en la Hoja de Ruta ICCVAM 2018.

En el contexto de este informe, el término MNE se refiere a cualquier tecnología, metodología, enfoque, o combinación de los mismos, que pueda utilizarse para proporcionar información sobre la evaluación de riesgos y peligros químicos, y que favorezca el reemplazo, reducción o

refinamiento (las 3R) del uso de animales. El presente informe se basa en los principios articulados en las directrices y documentos existentes (Tabla 1) aceptados por los Estados Unidos e internacionalmente, para abogar por un enfoque más flexible para generar confianza en las MNE. Aquí presentamos conceptos clave de validación, calificación y aceptación regulatoria tal como son aplicados en las MNE que incluyen pruebas en sistemas biológicos (por ejemplo, *in vitro*, ciertos *in quimico*, organismos modelo pequeños), reconociendo que pueden ser necesarias consideraciones adicionales para otros tipos de MNE, tales como predicciones mediante modelos computacionales (OCDE, 2007).

Tabla 1. Ejemplos de documentos de Estados Unidos e internacionales existentes, relacionados con la validación, calificación y uso regulatorio de MNE

Título del documento	Referencias
Validación y Aceptación Regulatoria de Métodos de Pruebas Toxicológicas (retirado a la fecha de publicación de este informe)	ICCVAM, 1997
Directrices del ICCVAM para la designación y presentación de métodos de prueba nuevos, revisados y alternativos	ICCVAM, 2003
Serie OCDE sobre Pruebas y Evaluaciones No. 34: Documento de Orientación sobre la Validación y Aceptación Internacional de Métodos de Prueba Nuevos o Actualizados para la Evaluación de Peligros	OCDE, 2005
Procedimientos recomendados en relación con la política de la CPSC sobre pruebas en animales	CPSC, 2012
Hoja de Ruta de Toxicología Predictiva de la FDA	FDA, 2017a
Calificación de herramientas para el desarrollo de dispositivos médicos: lineamientos para la industria, desarrolladores de herramientas y personal de la Administración de Alimentos y Medicamentos	FDA, 2017b
Plan Estratégico de la EPA para Promover el Desarrollo e Implementación de Métodos de Prueba Alternativos dentro del Programa TSCA	EPA, 2018
Hoja de Ruta Estratégica del ICCVAM para Establecer Nuevos Enfoques para Evaluar la Seguridad de Sustancias Químicas y Productos Médicos en los Estados Unidos	ICCVAM, 2018
Documento de Orientación sobre las Buenas Prácticas de Métodos In Vitro (GIVIMP, según sus siglas en inglés)	OCDE, 2018

Título del documento	Referencias
Lineamientos para la industria y desarrolladores de métodos de prueba: Evaluación del personal de la CPSC de métodos de prueba alternativos y enfoques de prueba integrados y datos generados a partir de dichos métodos para respaldar los requisitos de etiquetado de la FHSA	CPSC, 2020
Guía del proceso de calificación de herramientas de desarrollo de medicamentos, para la industria y el personal de la FDA	FDA, 2020
Plan de Trabajo de Métodos de Nuevo Enfoque de la EPA	EPA, 2021a
Plan estratégico de la EPA para reducir el uso de animales vertebrados en pruebas químicas	EPA, 2021b
Avance de las nuevas metodologías alternativas en la FDA	FDA, 2021a

2.0 Conceptos clave en la validación de MNE flexibles y aptas para el propósito pretendido

Los principios subyacentes de validación descritos en el GD 34 de la OCDE siguen siendo esenciales, pero los procesos utilizados para la validación deben permitir el desarrollo eficiente y oportuno de los MNE que sean aptos para su propósito, confiables y que proporcionen información relevante para las especies de interés. En concreto, El DG 34 (Documento de Orientación 34) de la OCDE establece que “los nuevos métodos de prueba se someten a validación para asegurar que emplean ciencia sólida y cumplen las necesidades regulatorias”, es decir, que los métodos son aptos para su propósito, lo que además de informar las decisiones regulatorias, podría incluir la selección y priorización de casos de uso. La guía también establece que “el proceso de validación debe ser flexible y adaptable” y que el desempeño debe “demostrarse utilizando una serie de productos químicos de referencia” y “evaluarse en relación con datos de toxicidad pertinentes existentes” (OCDE, 2005). Como tal, la confianza en las MNE no debe considerarse un estatus universal otorgado después de finalizar un proceso específico (por ejemplo, considerar que una NAM ya está “validada” después de completar con éxito un estudio de prueba en anillo); más bien, ganar confianza en las MNE debe considerarse como un proceso de evolución e iterativo, que requiere comunicación entre los desarrolladores de métodos, aquellos que toman decisiones regulatorias y las instituciones de validación (ICCVAM, 2018).

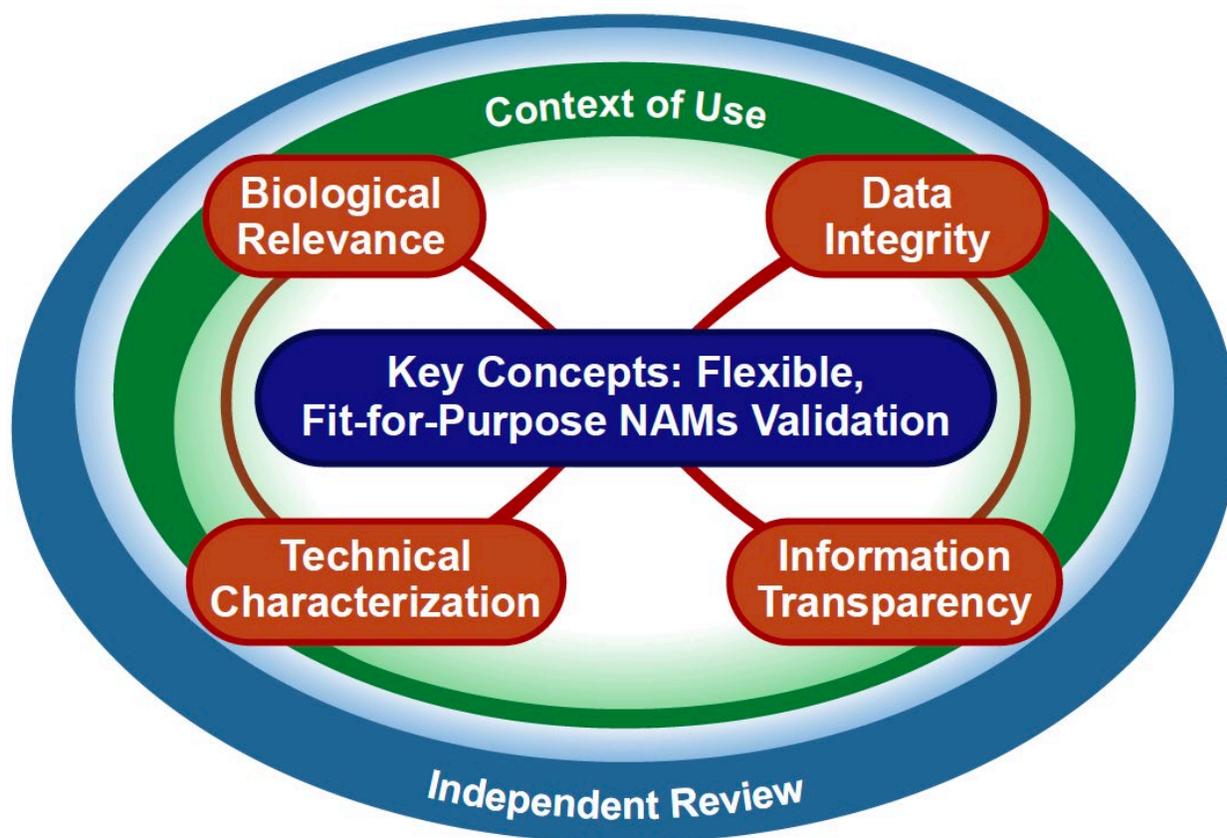


Figura 1. Conceptos clave a considerar durante el desarrollo e implementación de estrategias de validación de MNE flexibles y aptas para su propósito. Adaptado de van der Zalm et al. (2022).

Existen varios conceptos clave que es importante tener en cuenta al diseñar e implementar estrategias de validación flexibles y aptas para su propósito, representados en la Figura 1. Estos conceptos son ampliamente aplicables al uso de MNE para diversos fines, pero deben adaptarse, según sea necesario, a las circunstancias y aplicación específicas. Un concepto general a considerar es el CDU, o el propósito al que se destina la MNE (por ejemplo, detección, identificación de peligros, evaluación de potencia, punto de partida para la evaluación cuantitativa de riesgos, etc.). El CDU informará sobre la flexibilidad del proceso de validación y cómo se establece la aptitud de la MNE para un CDU específico. Los conceptos clave involucrados en ese proceso de validación flexible y apto para su propósito, son la importancia biológica, la caracterización técnica, la integridad de los datos y la transparencia de la información. Como se demuestra en la figura, estos aspectos están relacionados y son interdependientes. Por último, todas las partes del proceso de validación deberán estar sujetas a revisiones independientes. Estos conceptos clave son similares a los elementos esenciales propuestos por van der Zalm et al. (2022) en un marco para establecer la confianza científica en las MNE, pero aquí los estamos adaptando para que sean útiles en una gama más amplia de aplicaciones de las MNE. Cada uno de estos conceptos clave se detalla más en la Sección 3.0.

3.0 Aplicación de los conceptos clave para generar confianza en las MNE

La validación debe ser un proceso robusto pero flexible, en el que se establezca confianza científica al determinar la aptitud de una MNE para un CDU específico. Cuando sea posible y apropiado, debe existir evidencia que respalde que el uso de un método alternativo proporcionará información tan buena o mejor que el método existente y que hará que una revisión regulatoria conduzca a decisiones que sean igual de protectoras para la salud humana. El CDU determina los criterios detallados y la implementación de los conceptos clave. A fin de determinar si las MNE son aptas para un propósito particular, deben definirse con claridad tanto el CDU como la biología relevante que la MNE pretende abarcar. La MNE se deberá describir muy bien y proporcionar información técnicamente confiable que tenga importancia biológica y/o que proteja la salud para el punto final o el proceso en cuestión. La composición de la información proporcionada por la MNE deberá ser lo suficientemente transparente, como para permitir una revisión independiente que garantice la integridad y la fiabilidad. Cuando sea apropiado y posible, el generar confianza en las MNE puede incluir demostrar que éstas proporcionan información de calidad equivalente o mejor y relevancia para la especie de interés, para la toma de decisiones reglamentarias, en comparación, ya sea cuantitativa o cualitativa, con la información proporcionada por el método tradicional de ensayo en animales. Para evitar verse limitado por el status quo, deberá reconocerse la posibilidad de que la MNE pueda proporcionar información de mejor calidad y mayor importancia para la toma de decisiones regulatorias, que el método tradicional de ensayo en animales. Esto requiere un marco de validación que permita esta posibilidad y tenga en cuenta las circunstancias en las que podría no ser posible una comparación con los datos de métodos tradicionales de ensayo en animales.

3.1 Contexto de Uso

El establecimiento del CDU, o la utilización prevista de una MNE, incluye la elaboración de una declaración que describa completa y claramente la manera en que se pretende utilizar la MNE y su propósito regulatorio (si fuera aplicable). Las agencias federales de los Estados Unidos operan bajo estatutos y regulaciones particulares de cada agencia (véase, por ejemplo, Shaffer, 2021) y, por lo tanto, tienen diferentes criterios para que una MNE sea aceptable y aplicable a los requisitos individuales de cada agencia. En algunos casos, incluso dentro de una agencia (por ejemplo, diferentes centros de la Administración de Alimentos y Medicamentos de los Estados Unidos [FDA], u oficinas de la Agencia de Protección Ambiental de los Estados Unidos [EPA]) puede haber diferentes necesidades dependiendo del área de regulación. Esto se complica aún más a nivel internacional, ya que las regulaciones en otros países pueden diferir de aquellas de los Estados Unidos, de modo que los métodos aceptables en un país para un CDU en particular podrían no ser aceptables en otro lugar. En consecuencia, aunque una MNE puede validarse para abordar un criterio de valoración específico para uno (o varios) CDU bajo una directriz/regulación particular, podría no ser considerada aceptable bajo una directriz/regulación diferente, entre las diversas agencias, jurisdicciones reguladoras y países. Para comprender mejor este panorama regulatorio diverso, el ICCVAM ha recopilado información de una variedad de agencias, tanto a nivel nacional como internacional, para caracterizar las necesidades de las

agencias específicas para criterios de valoración de toxicidad particulares (NIEHS, 2023a; ejemplos mostrados en la Tabla 2).

Tabla 2. Manuscritos producidos por grupos de trabajo del ICCVAM, que proporcionan detalles sobre las necesidades de pruebas de la agencia para temas seleccionados

Las pruebas deben centrarse	Referencias
Pruebas de toxicidad aguda	Strickland y col., 2018
Ensayos de ecotoxicidad	Ceger y col., 2022
Extrapolación de <i>in vitro</i> a <i>in vivo</i>	Chang y col., 2022
Pruebas de nanomateriales usando MNE	Petersen y col., 2022a
Sensibilización de la piel	Daniel y col., 2018; Strickland y col., 2019
Pruebas de irritación de la piel y ojos	Choksi y col., 2019

Las MNE con diferentes CDU tales como detección/priorización e identificación de peligros, pueden ser diferentes de aquellas que están diseñadas para responder preguntas específicas referentes a mecanismos toxicológicos y/o apoyar evaluaciones cuantitativas de riesgos, y como tales, se pueden desarrollar y aplicar diferentes criterios para evaluar y validar las MNE para cada propósito. Muchos métodos nuevos son desarrollados en entornos académicos, para abordar cuestiones de investigación básica que pueden ser de naturaleza específica o exploratoria. Ocasionalmente, se reconoce que uno de estos métodos podría tener el potencial de desempeñar un papel distinto al de investigación básica en áreas más aplicadas, tales como la toma de decisiones regulatorias. Es importante que los desarrolladores de nuevos métodos se reúnan con agencias federales y/o usuarios previstos o partes interesadas, antes de validar una MNE, para garantizar que los estudios de validación diseñados para establecer confianza se adapten adecuadamente al CDU previsto.

Otras MNE pueden ser más útiles para el descubrimiento y desarrollo de productos. Se puede adoptar una MNE para detectar nuevas moléculas en busca de actividades deseables (por ejemplo, farmacológicas) o potencialmente indeseables (tóxicas). Estos datos pueden ser utilizados para tomar decisiones acerca de qué moléculas deben llevarse a un mayor desarrollo. La estandarización (normalización) de los métodos de prueba puede apoyar la innovación, al enfocar los esfuerzos y recursos en nuevos descubrimientos tecnológicos o terapéuticos, en vez de en el desarrollo o prueba de métodos. Estos usos de las MNE tienden a estar fuera de la autoridad reguladora de la mayoría de las agencias federales. No obstante, aún se recomienda aplicar los conceptos clave para establecer una confianza científica en las MNE, para asegurar que se estén obteniendo datos de alta calidad. Se reconoce que estos usos de las MNE pueden respaldar las 3R directamente y proporcionar datos preliminares y experiencia que puedan apoyar un mayor desarrollo de alguna MNE, permitiendo potencialmente su uso con fines regulatorios.

3.2 Relevancia biológica

La importancia de una MNE incluye la relación entre la prueba y el efecto en la especie objetivo y si el método de prueba es significativo y útil para un propósito definido, habiendo ya identificado las limitaciones. Una demostración adecuada de la importancia de una MNE contribuye de manera importante a la confianza en ella. La relevancia biológica puede demostrarse de diversas maneras, dependiendo de la información disponible, tal como se detalla a continuación.

Las consideraciones en torno a la relevancia biológica que se detallan en las siguientes subsecciones incluyen:

- ¿Qué tipo de información proporciona la MNE? ¿Existe una comprensión de la biología y los mecanismos que conducen al resultado o punto final?
- ¿Qué datos de referencia están disponibles para la evaluación comparativa del resultado que la MNE pretende cuestionar?
- ¿Cuáles son las consideraciones sobre si se deben comparar los resultados de una MNE con un método de laboratorio establecido y cómo hacerlo?

También pueden cobrar importancia consideraciones adicionales en torno a la relevancia biológica, dependiendo de las circunstancias específicas de aplicación de la validación.

3.2.1 Comprensión mecanicista

Es importante tomar en cuenta la biología de la especie de interés (generalmente humana, pero muchas veces otras especies) a la hora de evaluar la relevancia de una MNE. Las comparaciones de la información proporcionada por la MNE con la biología *in vivo* deberán estar respaldadas, siempre que sea posible, por el conocimiento mecanicista existente (por ejemplo, una ruta de resultado adverso [RRA] o un proceso biológico de importancia toxicológica). Anclar una MNE a una RRA establecida a través de un evento de iniciación molecular o uno o más eventos clave, puede ayudar a demostrar la relevancia biológica de la MNE. Por otro lado, la falta de una RRA establecida para el resultado que predice una MNE, no necesariamente excluye que la MNE sea potencialmente útil y dicha información de apoyo puede provenir de los conocimientos mecanicistas proporcionados por datos *in vivo* y de la comprensión de la biología de la especie objetivo. Es importante que una MNE sea caracterizada de manera integral y describa claramente qué evento biológico se está midiendo y cómo se relaciona con el resultado adverso o peligro preocupante.

Una descripción de la MNE debe abordar la plausibilidad biológica del modelo para predecir resultados *in vivo* y/o proporcionar un vínculo mecanicista con un proceso biológico, mecanismo o RRA. Por ejemplo, si una MNE va a predecir la posibilidad de que un producto farmacéutico de consumo humano induzca malformaciones fetales o letalidad embriofetal, debe haber una comprensión de los mecanismos de desarrollo embriofetal (por ejemplo, migración celular, diferenciación, vasculogénesis, neurulación, gastrulación) y su conexión con los efectos adversos del desarrollo posteriores estudiados con el modelo (FDA, 2021b). Si bien la relación con el

efecto *in vivo* que se predice puede ser de naturaleza correlativa, es más desafiante generar confianza en las predicciones de la MNE a partir únicamente de relaciones correlativas o empíricas; por lo tanto, se prefieren las pruebas con relevancia biológica clara para el proceso que se evalúa. La falta de comprensión de la relevancia biológica y mecanicista de una MNE puede limitar su aplicabilidad a los límites estrictamente definidos por los datos utilizados para validar la MNE y dificultar la extrapolación de las MNE a clases químicas fuera de las utilizadas para establecer y validar la MNE.

Una RRA es un marco organizativo útil para vincular perturbaciones moleculares y celulares con resultados adversos para la salud y puede usarse para desarrollar y anclar MNE que representan procesos toxicológicos importantes. Existen ejemplos de criterios de valoración que sí tienen RRA bien establecidas, tales como la sensibilización de la piel (Kleinstreuer et al., 2018) Para criterios de valoración que carecen de RRA bien establecidas, en su lugar se puede tomar en cuenta la relevancia mecanicista de la MNE, basándose en factores tales como rutas o procesos biológicos (Tabla 3).

Tabla 3. Ejemplos de criterios de valoración en los que se ha demostrado la relevancia biológica y mecanicista de las MNE*

Criterio de Valoración	Resumen	Referencias
Sensibilización de la piel	El criterio de valoración tiene una RRA en seres humanos relevante bien desarrollada, para la cual se mapean enfoques definidos que combinan varias MNE y se describen en la Directriz 497 de la OCDE.	Kleinstreuer y col., 2018; OCDE, 2021a
Alteraciones endocrinas	Hay disponibles modelos de rutas establecidas que utilizan MNE complementarias como parte de una estrategia integrada, para la actividad de receptores de estrógeno y andrógenos. La EPA acepta estas MNE para la detección de Nivel 1 en el Programa de Detección de Alteraciones Endocrinas.	Judson y col., 2015; Kleinstreuer y col., 2017; EPA, 2023
Neurotoxicidad del desarrollo	Existen RRA limitadas para este complejo criterio de valoración. En cambio, se ha desarrollado una serie de MNE que abarcan procesos críticos del neurodesarrollo humano. Se dispone de un DG de la OCDE sobre dicha serie que incluye estudios de caso de Enfoques Integrados para Pruebas y Evaluaciones (IATA).	Crofton y Mundy, 2021; OCDE, 2022a; OCDE, 2023

Criterio de Valoración	Resumen	Referencias
Toxicidad por inhalación	Se utilizó un enfoque alternativo empleando un ensayo <i>in vitro</i> basado en células humanas y modelado por computadora, para derivar un punto de partida para su uso en la evaluación de riesgos para la salud humana de la EPA. Este enfoque también fue publicado como un estudio de caso de IATA de la OCDE.	Corley y col., 2021; EPA, 2021c; OCDE, 2022b
Iritación ocular	Documento de orientación que describe un marco de pruebas alternativo para evaluar el potencial de irritación ocular de plaguicidas y productos plaguicidas. Se revisaron los métodos de prueba <i>in vivo</i> , <i>in vitro ex vivo</i> disponibles con respecto a su relevancia para la anatomía ocular humana y los mecanismos de toxicidad.	EPA, 2015; Clippinger y col., 2021
Iritación de la piel	Documento de orientación que propone un IATA para la corrosión e irritación de la piel, utilizando información existente, propiedades fisicoquímicas y otros métodos que no son de prueba.	OCDE, 2014

* Consulte la guía específica de la agencia para la aceptación de diferentes MNE.

De manera ideal, la descripción de la relación de la MNE con el efecto biológico de interés, se debe fundamentar en información disponible sobre la biología o mecanismo de acción relevante, para el criterio de valoración en cuestión en las especies de interés. En algunos casos, esto puede no ser posible y los datos de una especie diferente a la de interés pueden ser la única base disponible para comparación. No obstante, la fisiología de las especies de interés puede diferir de las especies de prueba sustitutas existentes, lo cual enfatiza aún más la necesidad de incorporar la comprensión mecanicista en la evaluación de la MNE. Por ejemplo, los aspectos anatómicos y fisiológicos del ojo de conejo difieren de los humanos (Clippinger et al., 2021) y los intervalos de susceptibilidad durante el desarrollo cerebral difieren entre especies (Smirnova et al., 2014; Tsuji y Crofton, 2012). En consecuencia, la relevancia biológica para las especies de interés y las consideraciones clave de exposición deben reconocerse al evaluar tanto la MNE como el método de prueba de referencia existente. Las evaluaciones de relevancia biológica y mecanicista podrían considerar, por ejemplo, la relevancia del tipo de célula utilizado, las características fisiológicas del órgano/tejido relevante bajo investigación, o la presencia de metabolitos de importancia para la especie asociados con la sustancia que se está probando (observando que los metabolitos encontrados en una especie podrían ser diferentes de los encontrados en otra o producidos en una MNE). Adicionalmente, se podría examinar la capacidad de un método para

evaluar un modo de acción o mecanismo en particular específico de una especie (Hartung, 2010; Madia et al., 2021; Parish et al., 2020).

Más allá de la traslación de especies, la variación entre individuos (por ejemplo, donantes de células o tejidos utilizados en ensayos *in vitro*) también puede ser subyacente a la variación biológica que puede afectar la respuesta medida, así como potencialmente agregar información adicional sobre la variabilidad entre individuos en una población (Harill, 2020). Si bien no se cubre a profundidad en el presente documento, los usuarios y desarrolladores de MNE deben ser conscientes de que la observación de un efecto (con respecto a la magnitud y en algunos casos, la presencia del efecto observado) así como el punto de partida o dosis a la que se produce el efecto, pueden ser modulados por la variación de secuencia genética presente en un panel de líneas celulares (Chiu et al., 2017). El uso de paneles de líneas celulares genéticamente diversas en pruebas basadas en células, puede ofrecer ventajas en la comprensión de la dinámica poblacional de la toxicidad y la respuesta fenotípica (Harill y McAllister, 2017), descubriendo variantes de secuencia genética que confieren susceptibilidad (Frick et al., 2015), lo cual permite la identificación de subpoblaciones singularmente susceptibles (Church et al., 2015), y facilita la medición de la variabilidad toxicocinética y toxicodinámica (Rusyn et al., 2022). Aunque a veces se utilizan líneas celulares derivadas de tumores en modelos *in vitro*, su composición genética podría no representar adecuadamente la variación genética en la población. Las recomendaciones específicas para incorporar la diversidad en las MNE dependerán del contexto. En consecuencia, el diseño de estudios adecuados para el propósito se debe desarrollar en cooperación con genetistas, para garantizar una cobertura adecuada de la variación de las especies objetivo, así como con bioeticistas y grupos de partes interesadas pertinentes si se va a cuestionar la variación genética humana (véase por ejemplo ICCVAM, 2022). Las recomendaciones específicas para evaluar la variación entre individuos están fuera del alcance de este informe y probablemente continuarán evolucionando con la ciencia emergente en este ámbito.

En algunos casos, las RRA complejas o los resultados que surgen de más de una RRA, pueden requerir la evaluación de múltiples aspectos del mecanismo del resultado adverso, para obtener una predicción adecuada del resultado global. Las MNE generalmente tienen limitaciones que no les permiten ser predictivos o técnicamente aplicables a todas las clases químicas. La MNE también puede estar limitada con respecto a la complejidad de la biología que pudiese representar. La combinación de más de una MNE con diferentes desempeños, aplicabilidad y coberturas biológicas en enfoques definidos (ED) puede mejorar su capacidad para predecir resultados (OCDE, 2017). Esto se puede lograr mediante el mapeo de MNE apropiadas a eventos clave a lo largo de una RRA, para asegurar una representación mecanicista suficiente para predecir el resultado adverso relevante para las especies de interés. Un ejemplo de mapeo de MNE a eventos clave a lo largo de una RRA que da como resultado un ED para abordar un criterio de valoración regulatorio, se encuentra en la “Directriz No. 497: Enfoques definidos acerca de la sensibilización de la piel” de la OCDE (OCDE, 2021a).

3.2.2 Compuestos de referencia

Los compuestos de referencia pueden tener una amplia gama de propósitos potenciales, que abarcan desde la ejecución de un ensayo hasta la evaluación de su calidad técnica y relevancia

biológica, tal como se muestra en la Tabla 4. Un compuesto de referencia podría funcionar bien para un propósito pero no ser adecuado para otro. Por ejemplo, un compuesto de referencia puede funcionar bien para un determinado método como control positivo (por ejemplo, un ensayo de neurotoxicidad del desarrollo), incluso si no diera como resultado el efecto adverso implicado en la especie objetivo (por ejemplo, debido a la incapacidad de atravesar la barrera hematoencefálica) y, por lo tanto, podría no ser adecuado como compuesto de referencia de criterio de valoración biológico. Dependiendo del CDU, los compuestos de referencia pueden incluir productos químicos de un solo componente, mezclas o extractos complejos (por ejemplo, suministrados por dispositivos médicos).

Tabla 4. Usos Seleccionados de Compuestos de Referencia

Conjuntos de Compuestos Discretos*	Propósito(s)	Criterios para la Selección de Compuestos
Control(es) Positivo(s)	Verifique si el método está funcionando como se esperaba.	Compuestos que se ha verificado que causan un cambio confiable, medible y estadísticamente significativo en una lectura de ensayo específico bajo las condiciones del método.
Compuestos de Desempeño	Evaluar la medida en que un método nuevo o modificado se compara con un método establecido similar - “yo también”.	Conjunto de compuestos que inducen de manera confiable una respuesta (o ninguna respuesta) en el método ya establecido.
Compuestos de Competencia	Evalúan el desempeño de un nuevo laboratorio utilizando un método establecido.	Compuestos que inducen de manera confiable una respuesta (o ninguna respuesta) en el método establecido; pueden ser un subconjunto de los compuestos de desempeño.
Compuestos de Referencia de Criterio de Valoración Biológica	<ol style="list-style-type: none"> 1) Evalúan la relevancia biológica de un método. 2) Comparan la concordancia de métodos diseñados para medir un resultado relevante para el mismo criterio de valoración <i>in vivo</i>, pero utilizando un contexto diferente (por ejemplo, una diferente especie, método o modalidad de medición). 	<ol style="list-style-type: none"> 1) Compuestos con evidencia de efecto <i>in vivo</i> (positivo) o ningún efecto (negativo) para el criterio de valoración de interés (idealmente del organismo objetivo de interés o de una especie comparativa adecuada). 2) Conjunto común de compuestos que han sido probados para el criterio de valoración de interés, en al menos un método confiable.

* Puede haber superposición de los compuestos elegibles en estos conjuntos.

Un aspecto clave para demostrar la validez científica de una MNE es evaluar su desempeño frente a métodos de prueba existentes y en uso, a menudo mediante el ensayo de compuestos de referencia de criterios de valoración biológicos, cuyas actividades biológicas están bien caracterizadas y son bien comprendidas. Los compuestos de referencia de criterio de valoración biológico generalmente incluyen compuestos que se ha demostrado que producen efectos adversos, de manera ideal con un rango de potencias, o que no producen tales efectos (Browne et al., 2019).

La identificación de compuestos de referencia de criterio de valoración biológico con efectos conocidos para resultados apicales particulares, puede ser un desafío. Esto es particularmente cierto en seres humanos, en donde a menudo faltan datos, ya que el campo de la toxicología se ha aplicado para prevenir la exposición humana a sustancias químicas potencialmente peligrosas. El éxito en este esfuerzo significa que se dispone de datos limitados en seres humanos, o que tales sustancias no se han caracterizado o medido adecuadamente en poblaciones humanas. Cuando se dispone de tales datos en seres humanos, pueden provenir de exposiciones accidentales o de bajo nivel, las cuales a menudo proporcionan datos cuantitativos limitados sobre los niveles de exposición o los resultados. Los estudios de caso pueden proporcionar indicaciones de peligro, pero con frecuencia no son adecuados para establecer correlaciones claras de causa-efecto.

A pesar de estos desafíos, se deben hacer esfuerzos para identificar compuestos con efectos en seres humanos relacionados con el resultado o mecanismo predicho por la MNE en desarrollo. Incluso si dichos datos son limitados en su capacidad para respaldar la evaluación cuantitativa del método, sí pueden proporcionar una verificación cualitativa sobre la importancia de la MNE. También puede ser difícil obtener datos epidemiológicos de alta calidad en seres humanos y alinearlos con datos toxicológicos; sin embargo, los datos epidemiológicos son útiles para identificar sustancias químicas ambientales con fuertes asociaciones con efectos adversos para la salud humana. Se han desarrollado marcos basados en evidencia para permitir la compilación de grandes cantidades de información, seguida por una selección transparente y objetiva de conjuntos de datos más estrechos e importantes (Wikoff et al., 2020), mejorando así la accesibilidad de los datos. Estos marcos incluyen mapas sistemáticos de evidencia, que proporcionan una visión general amplia de una base de evidencia, y revisiones sistemáticas posteriores, que proporcionan una evaluación más estrecha y completa de una cuestión en investigación en particular, permitiendo una integración más rápida de los datos epidemiológicos en seres humanos con otras fuentes de información existentes (Wolffe et al., 2019). En combinación con líneas mecanicistas de evidencia y datos en animales heredados, estos datos podrían usarse para definir compuestos de referencia de criterio de valoración biológico, para ayudar a establecer confianza científica en las MNE (ver Krishna et al., 2021 para un ejemplo).

Siempre que sea posible, los compuestos que se ha demostrado que causan efecto(s) relevante(s) en las especies de interés, deben utilizarse como controles positivos durante la evaluación de la robustez y relevancia de una MNE. Petersen (2021) ha descrito características a considerar al elegir controles positivos (por ejemplo, sustancias químicas probadas utilizando medidas de control en proceso cada vez que se ejecuta el ensayo) y muchas de las mismas consideraciones serán relevantes para compuestos de referencia que evalúan una MNE. Deben incluirse compuestos de referencia de criterio de valoración biológico tanto positivo como negativo,

compuestos de desempeño y compuestos de competencia, y deberán ser seleccionados con el mismo cuidado. Debido al sesgo histórico en la literatura científica que enfatiza los resultados positivos (reconociendo que no todos los positivos reportados en la literatura son verdaderos positivos), los compuestos de referencia confiables que son negativos para el criterio de valoración objetivo, pueden ser más difíciles de identificar. Los esfuerzos de curación de datos centrados en resultados negativos deberán ser tomados muy en cuenta para su publicación, y los esfuerzos que prueben compuestos hasta concentraciones relativamente altas en una amplia gama de objetivos, pueden proporcionar información valiosa para ayudar a resolver este problema.

Es posible que no todos los CDU sean abarcados por una lista definitiva, y la selección de compuestos de referencia debe tener en cuenta las necesidades regulatorias particulares de ese CDU. En algunos casos, también puede ser apropiado definir subconjuntos de compuestos de referencia de criterio de valoración biológico con respecto al mecanismo que se está evaluando (en lugar del criterio de valoración o punto final apical) para cada NAM dentro de una batería. Por ejemplo, pocos compuestos califican como compuestos de referencia para la neurotoxicidad del desarrollo y, en su lugar, cada proceso evaluado en la batería (por ejemplo, proliferación o excreción de neuritas) tiene sus propios compuestos de referencia de criterio de valoración biológico específicos del ensayo. En consecuencia, un compuesto de referencia de criterio de valoración biológico puede ser positivo en un ensayo, pero negativo en otro ensayo que esté evaluando diferentes mecanismos o procesos.

Se requieren listas de compuestos de referencia (por ejemplo, para evaluar la capacidad de transferencia) en el proceso de validación formal delineado por la OCDE. Sin embargo, la compilación de compuestos de referencia confiables requiere mucho tiempo y muchos recursos. Los esfuerzos para facilitar el rápido desarrollo de listas curadas de compuestos de referencia, mediante revisiones sistemáticas rigurosas y procesos de automatización cuando es factible, han permitido una evaluación más robusta de la sensibilidad y especificidad de las MNE (Judson et al., 2019; Thomas et al., 2019). Un ejemplo de una lista de compuestos de referencia de criterio de valoración biológico para la calificación de MNE, se puede encontrar en la revisión de 2021 de la FDA del Consejo Internacional para la Armonización (ICH) de Requisitos Técnicos para Productos Farmacéuticos de Consumo Humano, S5 (R3) Detección de Toxicidad Reproductiva y del Desarrollo para la Guía de Productos Farmacéuticos de Consumo Humano para la Industria (FDA, 2021b). La lista incluye compuestos con resultados positivos y negativos. Dicha guía señala que estos compuestos, así como otros, pueden usarse para respaldar la calificación de un ensayo alternativo o batería de ensayos para CDU particulares.

3.2.3 Comparación con Métodos Existentes en Animales de Laboratorio

La norma para establecer confianza científica en una MNE y obtener aceptación regulatoria ha incluido generalmente la consideración de si la MNE puede proporcionar información sobre una utilidad equivalente o mejor, calidad científica y/o relevancia, que el método de prueba existente utilizado para la toma de decisiones regulatorias (según corresponda dentro del marco regulatorio de cada organismo). Por ejemplo, un criterio en el DG 34 de la OCDE para validar cualquier nuevo método de prueba es: “el método genera datos para fines de evaluación de riesgos que son al menos tan útiles como, y preferiblemente mejores que, los obtenidos utilizando métodos

existentes. Esto dará un nivel comparable o mejor de protección para la salud humana o el medio ambiente” (OCDE, 2005). La Ley de Control de Sustancias Tóxicas de los Estados Unidos enmendada (sección 4(h)(1)(B)) incluye consideraciones específicas para las MNE y ordena que la EPA fomente y facilite el “uso de métodos y estrategias de prueba científicamente válidos, que reduzcan o reemplacen el uso de animales vertebrados, a la vez que proporcionen información de calidad científica equivalente o mejor y relevancia, que respaldarán decisiones regulatorias” (15 USC §2601, 2016).

Históricamente, el concepto de “equivalente o mejor” se ha basado en una comparación directa con los tradicionales datos de pruebas en animales. No obstante, para lograr la meta de “mejor” información, debe reconocerse que las MNE podrían no proporcionar la misma información generada por el método tradicional de prueba en animales y los resultados de la MNE podrían no alinearse directamente con los resultados de la tradicional prueba en animales (para ejemplos, véanse Clippinger et al., 2021; Hoffmann et al., 2018; ICCVAM, 2018; Kolle et al., 2017; Petersen et al., 2022b; Piersma et al., 2018; Prior et al., 2019; Sewell et al., 2017). En algunos casos, la MNE puede proporcionar información biológicamente relevante, conocimientos mecanicistas o criterios de valoración suficientemente sensibles, adecuados para el proceso de toma de decisiones regulatorias y puede que no sea necesaria una comparación con los datos de los tradicionales métodos de ensayo en animales. Por ejemplo, algunos objetivos farmacológicos o toxicológicos podrían no existir en especies no humanas, por lo que los estudios en animales podrían no ser relevantes para evaluar los posibles efectos en seres humanos mediados por dichos objetivos. Las MNE basadas en seres humanos podrían proporcionar información biológicamente relevante en estos casos. Además, las MNE a menudo proporcionan información mecanicista en lugar de datos sobre los criterios de valoración apicales medidos con métodos de prueba en animales (por ejemplo, mientras que una reducción observada del peso corporal en un animal podría no dilucidar el mecanismo de toxicidad subyacente, las MNE podrían ser capaces de proporcionar estos conocimientos mecanicistas). También hay circunstancias en las que el modelo animal podría estar midiendo un criterio de valoración biológico complejo que es relevante para el CDU, pero que no está adecuadamente cubierto por la MNE en cuestión. En estos casos, una comparación con los métodos tradicionales de prueba en animales puede ser la opción más conveniente, en particular cuando el mecanismo no se entiende por completo. El objetivo de la evaluación de la MNE es demostrar que ésta proporciona información que conduce a una decisión regulatoria similar a la que se tomaría con base en los métodos existentes.

Las comparaciones entre las MNE y los métodos existentes en animales de laboratorio, según corresponda, deben considerar la confiabilidad y reproducibilidad de los métodos de prueba de referencia en animales, incluida la comprensión de las razones de la varianza observada en ambos tipos de métodos, si es posible. Cuando están disponibles, el uso de datos de referencia de las especies de interés permite la evaluación de una MNE frente a la respuesta relevante para la especie. Si se dispone de datos de tiempo-curso *in vivo*, se pueden usar MNE tales como sistemas microfisiológicos, para simular la toxicocinética y luego la extrapolación por computadora del sistema microfisiológico puede ser comparada con los datos provenientes de animales o seres humanos. El modelado farmacocinético fundamentado fisiológicamente permite la estimación de concentraciones internas a partir de dosis de referencia y la comparación con concentraciones de actividad *in vitro*. Estos modelos también pueden aplicarse en un enfoque de dosimetría inversa,

para realizar una extrapolación de *in vitro a in vivo* (EIVIV) y predecir dosis administradas equivalentes que resultarían en concentraciones plasmáticas, o en el tejido objetivo, en donde se observa bioactividad en la MNE. Cuando los seres humanos son la especie de interés, la evidencia de alta calidad epidemiológica, clínica o por observación, de los efectos, puede ser útil para proporcionar información sobre la generación de confianza científica para la MNE. Sin embargo, esos datos de referencia en seres humanos rara vez están disponibles para la mayoría de los criterios de valoración y productos químicos, y las comparaciones a menudo deben fundamentarse en datos de una especie diferente. Los datos de estudios en animales se pueden curar y comparar para obtener listas de estándares de referencia con resultados reproducibles, robustos y relevantes. En algunos contextos, como las aplicaciones ecológicas, la homología de secuencia de objetivos moleculares o la conservación de mecanismos biológicos, pueden ser útiles para apoyar una comparación con datos provenientes de diferentes especies (Farmahin et al., 2013; LAlone et al., 2016).

Diversas publicaciones han evaluado los resultados de métodos de prueba de referencia basados en animales, para una variedad de criterios de valoración y han demostrado que los resultados de estas pruebas demuestran un grado variable de reproducibilidad (Browne et al., 2018; Dumont et al., 2016; Karmaus et al., 2022; Kleinstreuer et al., 2018; Luechtefeld et al., 2016; Pham et al., 2020; Rooney et al., 2021). La variabilidad observada podría reflejar, por ejemplo, factores biológicos inherentes, errores al reportar, diferencias de protocolo o variaciones en la pureza de la sustancia de prueba. Siempre que sea posible, deben explicarse los resultados contradictorios entre el método de prueba en animales de referencia y los datos de la MNE, incluyendo hacer referencia a la biología de la especie de interés. Esto es necesario para proporcionar un contexto realista sobre las capacidades de los datos de animales de laboratorio, y por lo tanto, establecer expectativas adecuadas en torno a la capacidad máxima de desempeño de las MNE que se comparan con el método de prueba de referencia (Browne et al., 2019).

El valor relativo de comparar las MNE con las pruebas en animales heredadas y la relevancia biológica y mecanicista para las especies de interés de las MNE, deben considerarse con base en la calidad de los datos disponibles. Siendo realistas, en muchos casos aquellos que buscan establecer confianza en una MNE necesitan considerar el uso histórico de estudios en animales y el requisito de comparación con los métodos existentes como una importante línea de evidencia. Bajo muchos requisitos legales y reglamentarios, la información derivada de las MNE no debe ser menos protectora que los métodos existentes. Idealmente, el método será más predictivo y podría permitir una generación más rápida y completa de datos relevantes en muchos productos químicos, en donde de otro modo los datos podrían ser escasos o ausentes, generando de este modo confianza entre los reguladores y las comunidades de interesados. Puede haber CDU específicos que permitan excepciones a esta regla; sin embargo, la aceptación de tales MNE como métodos calificados estará sujeta a los requisitos específicos de la institución u organismo regulador correspondiente. El alcance de dichas comparaciones puede variar dependiendo de la cantidad y calidad de los datos disponibles y de la profundidad de comprensión de la relevancia para la especie de interés, tanto de la MNE como del método de referencia. Cuando el desempeño de la MNE se evalúa con base en el método de prueba en animales de referencia, la cantidad de varianza aceptable en la MNE debe considerarse en relación con a la varianza observada en los datos *in vivo* y en función del CDU previsto.

3.3 Caracterización Técnica

La caracterización técnica es un componente clave del desarrollo de MNE para uso generalizado. Incluye una evaluación de las fuentes de variabilidad en la MNE, diseñar el ensayo para incluir medidas de control relevantes, evaluar el rango de sustancias de prueba para las cuales se puede utilizar el ensayo (es decir, el dominio de aplicabilidad), y desarrollar un enfoque adecuado de análisis estadístico de los datos. Un aspecto clave para demostrar la validez científica de una MNE, es que el ensayo esté suficientemente bien caracterizado desde una perspectiva técnica para asegurar que sea robusto, confiable y reproducible. La MNE incluye tanto el sistema de prueba mismo, como el método de cuantificación del criterio de valoración. Existen aspectos de caracterización técnica que se solapan con conceptos clave previos, tales como la relevancia biológica, lo que informa la selección de compuestos de referencia para diferentes propósitos (Tabla 4). Ya se ha presentado (en la Sección 3.2.2) el examen de desempeño de una MNE frente a un conjunto bien definido de compuestos de referencia de criterio de valoración biológico con bioactividades relevantes, por lo que esta sección se enfocará en aspectos técnicos tales como las herramientas de calidad (Ishikawa, 1985), el desarrollo de métodos, la documentación y normas.

3.3.1 Incorporación de Herramientas de Calidad Seleccionadas

En general, la caracterización técnica de las MNE encaja en un marco con pasos generales (Figura 2): 1) examen inicial de la relevancia científica de la MNE, 2) evaluación conceptual, 3) evaluación en laboratorio, 4) análisis estadístico de datos y presentación de informes, y 5) evaluación interlaboratorios (si fuera necesario).¹ Los resultados de los pasos 2 a 5 están interrelacionados, y los resultados de cada paso pueden tener impacto en todos los demás. Por ejemplo, los resultados de las pruebas interlaboratorios podrían revelar que se deben realizar pruebas adicionales en laboratorio o que es necesario revisar el protocolo para incluir una nueva medición de control.

¹Podría haber ejemplos de métodos para los que la evaluación interlaboratorios no es necesaria o no es factible, tales como ensayos de detección robotizados de alto rendimiento a gran escala y enfoques *in silico* basados en aprendizaje automático.

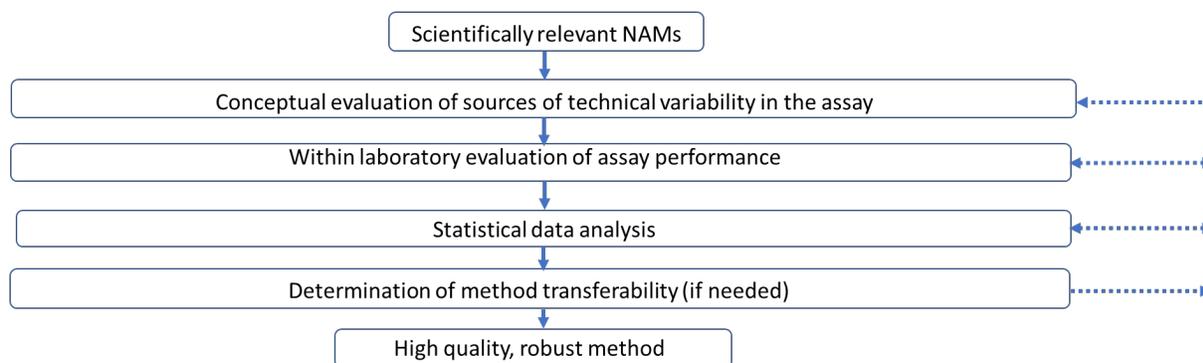


Figura 2. Entorno para el desarrollo de MNE robustas. Las líneas continuas indican los pasos que deben seguirse en el orden sugerido. Las líneas punteadas indican una dirección que se puede tomar, si es necesario, para reevaluar cualquiera de los pasos anteriores. Adaptado y reimpresso con permiso de Petersen et al. (2022b).

El paso de examen inicial incluye una evaluación de la relevancia biológica de un ensayo, su potencial CDU y el potencial de alcanzar un nivel suficiente de calidad técnica. Si la MNE carece de un CDU claro o de relevancia biológica o presenta importantes problemas técnicos de calidad que potencialmente no pueden resolverse, es posible que la MNE no sea adecuada para uso regulatorio. En la fase de evaluación conceptual, se revisa la MNE para evaluar las fuentes esperadas de variabilidad y diseñar medidas de control. En la evaluación intralaboratorio, se pueden realizar experimentos para evaluar la robustez del ensayo, establecer un rango típico para las medidas de control (por ejemplo, controles negativos, vehiculares y positivos), e identificar si existen interacciones entre las medidas de control y el resultado de la prueba. La información de la evaluación conceptual y la evaluación intralaboratorio permiten diseñar un modelo estadístico para evaluar el resultado medio del ensayo, caracterizar la variabilidad y establecer criterios para una respuesta positiva o negativa y la confianza estadística de esta determinación.

Por último, puede ser necesario que algunas MNE sean sometidas a una evaluación de transferibilidad mediante pruebas interlaboratorios. Las pruebas interlaboratorios pueden revelar potencialmente los pasos de un protocolo que se interpretan de manera diferente entre los laboratorios y las revisiones que se pueden hacer para mejorar la calidad técnica de la MNE. Los estudios de transferibilidad también pueden identificar problemas que surgen de diferencias en la implementación técnica o ejecución de un protocolo (por ejemplo, pipeteo en un laboratorio versus automatización en otro) que necesitan cambios.

En la Tabla 5 se listan varias herramientas de calidad que pueden ser útiles en la caracterización técnica de una MNE. Un tratamiento más detallado de este tema está disponible en Petersen et al. (2022b). Parte de la información generada a partir de estas herramientas puede permanecer interna con el desarrollador del método de prueba, mientras que otra información puede ser solicitada para apoyar la revisión de una agencia.

Tabla 5. Herramientas de Calidad Utilizadas con Frecuencia en la Caracterización Técnica de las MNE*

Herramienta de calidad	Descripción	Beneficio/Utilidad
Diagramas de flujo	Diagrama de cada paso en un protocolo.	Optimizar la identificación y cobertura de los pasos del protocolo que pueden ser monitoreados por medidas de control.
Análisis de Causa y Efecto	Revisar literatura y antecedentes del ensayo; hacer diagrama de todas las fuentes esperadas de variabilidad.	Identifica fuentes clave de variabilidad y aspectos de un método que pueden ser difíciles de estandarizar. Orienta las pruebas de robustez y la selección de medidas de control.
Gráficos de Control	Medidas de control para evaluar el desempeño técnico, incluyendo experimentos preliminares de ejecución única, mediciones periódicas a una frecuencia predeterminada y medidas de control en proceso realizadas cada vez que se ejecuta el ensayo. Los gráficos de control monitorean las medidas de control a lo largo del tiempo.	Prueban posibles sesgos, evalúan el rendimiento y la calibración del instrumento y miden las fuentes clave de variabilidad a lo largo del tiempo y entre experimentos.
Hojas de Verificación	Registran datos clave, metadatos y medidas de control.	Monitorean las medidas de control en el proceso y sirven de respaldo para la resolución de problemas cuando surjan, apoyan el análisis de datos y la reproducibilidad.
Gráficas de Dispersión	Grafican todos los datos de las medidas de control y los resultados de la sustancia de prueba.	Evalúan si existe una interacción entre diferentes medidas de control en proceso o entre estas medidas de control y los resultados de la sustancia de prueba.

*Consulte detalles y ejemplos en el Apéndice B.

3.3.2 Mejores Prácticas de Control de Calidad

En esta sección se describen los factores técnicos que las agencias federales pueden considerar como mejores prácticas para la evaluación de MNE, incluyendo los ED y los Enfoques Integrados para Pruebas y Evaluaciones (IATA, Integrated Approaches to Testing and Assessment). Los factores listados aquí pueden aplicarse o no a todas las MNE. Además, otros factores que no figuran en la guía, como información adicional o pruebas, pueden ser necesarios para una agencia al evaluar la MNE. Se alienta a los desarrolladores y/o patrocinadores a comunicarse directamente con la agencia federal a la que pretenden presentar métodos o datos, para determinar factores adicionales que pueda ser necesario revisar.

La presentación de una MNE para su evaluación reglamentaria debe incluir una descripción de cualquier estudio intralaboratorio o interlaboratorios, si es que fueron realizados. Debe precisarse si la MNE se comparó con un método de referencia: otra MNE, datos *in vivo* en animales, o datos en seres humanos. También puede ser necesaria información sobre el método de referencia utilizado para la comparación (es decir, el conjunto de datos de referencia) durante la evaluación. La evaluación de una MNE se puede lograr a través de varios procesos diferentes e implica documentación, uso de investigaciones de laboratorio específicas, que las características de desempeño de un método sean adecuadas y confiables para la(s) aplicación(es) prevista(s). La aceptabilidad de los datos se relaciona directamente con los criterios utilizados para evaluar el método.

Además de servir para documentar el desempeño y caracterizar la aplicabilidad de la MNE, las buenas prácticas científicas, técnicas y de calidad aseguran que el proceso de evaluación general sea eficiente y efectivo, lo que conduce a una mayor confianza en el método propuesto. Los desarrolladores deben conservar toda la información necesaria para calibrar, operar y mantener cualquier equipo, tal como manuales de equipo y software, certificados y garantías de confirmación de calidad y seguridad, documentación de versiones de software, algoritmos de modelado, bases de datos curadas y materiales de capacitación. Los desarrolladores también deben mantener documentación de los proveedores de materiales, células y reactivos, siempre que esta información sea relevante para la evaluación de una MNE (discutido en detalle en la Sección 3.3.3). Para obtener más detalles sobre prácticas de calidad, procedimientos de equipos y documentación a retener, consulte el Documento de Orientación de la OCDE sobre Buenas Prácticas de Métodos In Vitro (GIVIMP; OECD, 2018). Se recomienda un informe de Calidad de Instalación/Calidad de Operación/Calidad de Desempeño (IQ/OQ/PQ, según sus siglas en inglés) para cada instrumento utilizado y constituye un requerimiento si se pretende que los estudios cumplan con las Buenas Prácticas de Laboratorio (BPL).

3.3.2.1 Información Relevante para los Métodos en Células/Tejidos

Debe mantenerse la documentación del origen de las células y tejidos empleados en cualquier método de prueba. La información retenida para revisión, de manera ideal debería incluir (pero no limitarse a):

- Consideraciones éticas, legales y de seguridad

- Especie/raza/sexo
- Información demográfica (si es de importancia)
- Fuente/proveedor
- Número de donantes
- Órgano/tejido de origen
- Tipo(s) de célula(s) aisladas
- Secuencias de objetivos moleculares manipulados por ingeniería genética, cuando corresponda
- Técnica y fecha de aislamiento
- Clasificación de bioseguridad
- Identificación y autenticación de líneas celulares
- Pruebas serológicas para agentes infecciosos (por ejemplo, pruebas de micoplasma)
- Número de pases/duplicación de población de cultivo celular
- Tiempo de duplicación
- Estabilidad genética
- Información de expresión genética y proteica (si es relevante)
- Tratamiento previo

Para obtener información más detallada, consulte el DG sobre Buenas Prácticas de Cultivo Celular y Tisular 2.0 (Pamies et al., 2022) y el documento GIVMP de la OCDE (OECD, 2018).

Se debe proporcionar documentación sobre la cuarentena de nuevas células y tejidos bajo condiciones de almacenamiento adecuadas en el laboratorio antes de su uso en pruebas. Los ensayos que utilizan células o tejidos vivos deben incluir un ensayo de viabilidad celular.

Todos los laboratorios participantes deben mantener los registros y documentación de desempeño de todo el equipo de laboratorio (lectores de placas, incubadoras, refrigeradores/congeladores, etc.)

El registro de células en bancos celulares puede contribuir a la rastreabilidad y disponibilidad a largo plazo. Los bancos celulares deben incluir documentación sobre la densidad celular y el número de pases de las células. Se deben mantener registros de preparación de reactivos (por ejemplo, utilizando hojas de verificación para rastrear cualquier reactivo o consumible utilizado). Esto también es relevante para métodos *in vitro* que no utilizan células/tejidos. Algunos ejemplos son (pero no se limitan a):

- Proveedor
- Número de catálogo
- Número de partida/lote
- Fechas de preparación
- Fechas de caducidad
- Nombre del analista

El patrocinador o desarrollador del estudio debe proporcionar toda la información de seguridad en lo que respecta al método o metodología que se está desarrollando, así como todas las regulaciones referentes al uso, transporte y eliminación de todos los materiales peligrosos.

3.3.2.2 Evaluación del Método Analítico Utilizado en la MNE

El desarrollador deberá proporcionar datos que demuestren claramente que la sustancia detectada o cuantificada es la sustancia química o analito pretendido de interés. En la Tabla 6 se enumeran y describen las mejores prácticas de control de calidad relacionadas con los métodos analíticos.

Tabla 6. Evaluación de Métodos Analíticos*

Evaluación del Método Analítico	Descripción	Beneficio/Utilidad
Límites de Detección y Cuantificación	Es la menor cantidad o concentración de un analito que puede detectarse o cuantificarse de manera confiable, por encima del blanco de reactivos o dentro de la curva estándar. La concentración más alta en la curva estándar determina el límite de cuantificación superior.	Establece el rango de un método analítico.
Identificación de Interferencias	Identifica cuándo los componentes del método alteran falsamente la señal detectada.	Determina si hay interacciones entre los componentes del método y evita el reporte de resultados artificiales.
Evaluación de la Precisión Analítica	Caracteriza la precisión del método analítico utilizado y cualesquiera otras pruebas de precisión, tales como aquellas que evalúan la variabilidad del desempeño cuando diferentes personas utilizan el método propuesto o cuando se emplea instrumentación diferente para el método. Esto puede ser evaluado mediante estudios de comparación interlaboratorios.	Genera confianza en la confiabilidad del método analítico y evalúa fuentes de variabilidad de laboratorio.

Evaluación del Método Analítico	Descripción	Beneficio/Utilidad
Estabilidad de los Materiales Utilizados en las MNE	Es la capacidad de los materiales utilizados en la MNE (por ejemplo, sustancias de prueba, aparatos de prueba, sistema de exposición, reactivos y analitos) para producir resultados similares y aceptables durante un período de tiempo, en un ambiente determinado.	Caracteriza los materiales del método y asegura que se obtengan datos confiables de manera consistente para un método en particular.
Pruebas de Robustez	Prueba la capacidad de un método para ser reproducido bajo diferentes condiciones o circunstancias, sin que se produzcan diferencias inesperadas en los resultados obtenidos.	Determina el rango de parámetros en los que el ensayo funciona de manera aceptable.
Análisis de Recuperación	Prueba un método de extracción mediante la comparación de los resultados de las muestras extraídas con los de muestras sembradas en una matriz similar y/o en blancos sembrados.	Verifica la eficiencia y reproducibilidad de un método de extracción.
Análisis Técnico del Dominio de Aplicabilidad	Obtiene datos adecuados del método de prueba para sustancias químicas y/o productos representativos de aquellos relevantes para el CDU específico para el que se propone la prueba y describe con claridad las propiedades fisicoquímicas del dominio de aplicabilidad.	Reduce la incertidumbre con respecto al desempeño del ensayo para su uso con diferentes sustancias y/o productos químicos, y proporciona métodos y criterios para determinar cuándo una sustancia química se encuentra dentro del dominio de aplicabilidad.

Evaluación del Método Analítico	Descripción	Beneficio/Utilidad
Control Positivo	Identifica el o los compuestos de control positivo relevantes para el criterio de valoración y dentro del intervalo de detección del ensayo.	Proporciona una base consistente y confiable para la comparación de los resultados de la sustancia de prueba.
Estándares de Referencia para la Calibración de Instrumentos	Se utilizan para calibrar instrumentos usando estándares de calibración y/o muestras de control de calidad.	Asegura mediciones confiables e identifica posibles fuentes de incertidumbre.
Configuración de Especificaciones	Establece las especificaciones para medidas de control en proceso basadas en resultados de pruebas intralaboratorio y/o interlaboratorios, mediante el uso de un enfoque estadístico.	Asegura criterios suficientemente estrictos de medidas de control, que conducen a resultados de prueba robustos de la sustancia de prueba y carencia de sesgos.

*Puede consultar detalles y ejemplos en el Apéndice C.

3.3.2.3 Evaluación de la Exactitud y/o Concordancia de la MNE con los Estándares de Desempeño

La exactitud y concordancia son muy similares en definición y a menudo se consideran intercambiables, dependiendo del contexto en un documento o pasaje determinado. La concordancia a menudo se define como la comparación de dos métodos o pruebas fundamentadas en los resultados obtenidos, mientras que la exactitud a menudo se define como la comparación de un método o prueba con un método o resultado de prueba de referencia. En la evaluación estadística de métodos de prueba y datos asociados, a menudo se utiliza una evaluación de exactitud y/o concordancia. En este contexto, la exactitud se define como la proporción de predicciones correctas entre el número total de resultados. Otros parámetros estadísticos utilizados al discutir la exactitud o concordancia de métodos incluyen la sensibilidad, especificidad, predictividad positiva y negativa, y la tasa de falsos positivos y falsos negativos.

Siempre que sea posible, se pueden utilizar normas de desempeño bien definidas (por ejemplo, un conjunto equilibrado de sustancias de referencia que se sabe producen resultados positivos y negativos) para verificar la respuesta y la validez del método. Se deben reportar las medidas cuantitativas de concordancia (es decir, la sensibilidad, especificidad, predictividad positiva y negativa, y la tasa de falsos positivos y negativos). Al comparar un método de prueba propuesto con un método con normas de desempeño establecidas (por ejemplo, métodos tipo “yo también”

para las directrices de prueba de la OCDE [TG]) que produce datos funcional y mecánicamente similares, la concordancia (incluyendo los datos discordantes) de ambos métodos deberá evaluarse uno frente al otro y contra un método de referencia.

3.3.2.4 Procedimientos Normalizados de Operación y Detalles del Método

Se recomienda que el método de prueba propuesto cuente con procedimientos Normalizados de Operación (PNO) bien documentados, para respaldar el desempeño consistente del método y las actividades de laboratorio relacionadas, tales como el manejo del sistema de prueba y la calibración del equipo. Los PNO deben abarcar todos los aspectos de las pruebas y análisis. Los PNO deben incluir:

- Sistemas de rendición de cuentas que aseguren la integridad de los artículos de prueba (por ejemplo, mantenimiento de registros, seguridad y cadena de custodia de las muestras).
- Preparación de muestras y herramientas analíticas, tales como métodos, reactivos (incluyendo, cuando corresponda, el fabricante, número de catálogo, número de lote, etc.), equipo e instrumentación.
- Procedimientos de control de calidad y verificación de resultados.
- Detalles del método, incluyendo la descripción y formulación completa del producto, el sistema de exposición, el volumen/peso/solubilidad/protocolo de dosificación de la sustancia de prueba y el rango de exposición/dosis apropiado.

La parte que presenta la solicitud también debe proporcionar una lista de características y criterios operativos, para evaluar el desempeño de la prueba y los resultados. La información operativa y los criterios para los sistemas técnicos que integran la MNE pueden variar, pero los criterios podrían incluir gráficas de control de calidad u otros estándares de desempeño para todos los controles, estándares, exposiciones (dosis y duración) y grupos experimentales. Debe incluirse una descripción de los métodos estadísticos utilizados para evaluar los datos. Además, los desarrolladores deben describir cómo se evaluaron la incertidumbre experimental, incertidumbre estadística, interferencia y los antecedentes.

3.3.3 Documentación

En esta sección se describen las mejores prácticas para la documentación de las MNE (incluidos EDs e IATA). Los factores aquí listados pueden aplicarse o no a todas las MNE. Además, una agencia podría necesitar documentación adicional para evaluar la MNE. Se alienta a los desarrolladores y/o patrocinadores a comunicarse directamente con la agencia federal a la que pretenden presentar métodos o datos, para determinar factores adicionales que pudiera ser necesario revisar. Para algunas instituciones, es posible que pudiera ser necesaria una revisión por colegas independientes de la MNE y datos asociados, antes de la revisión por parte de dichas instituciones.

Para los nuevos métodos, la documentación debe incluir una descripción del método de prueba propuesto y cómo puede ser relevante para fines reglamentarios o cómo encajaría en un CDU

específico. La información relevante incluiría (pero no se limitaría a) cualquier información mecanicista y de relevancia biológica del método de prueba y cualquier CDU propuesto (por ejemplo, aplicada a la evaluación de riesgos). En la documentación se debe incluir el dominio de aplicabilidad de taxones humanos o apropiados. A continuación se detallan otros aspectos importantes para la documentación.

Se dispone de recursos adicionales que proporcionan orientación sobre la documentación. Para los métodos *in vitro*, los desarrolladores pueden consultar el documento GIVMP de la OCDE (OECD, 2018).

3.3.3.1 Identidad y Pureza de la Sustancia de Prueba

Para todas las sustancias probadas en la MNE (por ejemplo, controles, compuestos de referencia, otras sustancias de prueba), como mínimo, se debe reportar la identidad de la(s) sustancia(s), idealmente un identificador único (por ejemplo, CASRN, SMILES, InChIKey) e información sobre la pureza de la sustancia proporcionada por el proveedor. Si el laboratorio cuenta con las capacidades necesarias, se agrega beneficio al realizar medidas de control de calidad adicionales empleando métodos analíticos para evaluar la pureza e identidad de la sustancia. La capacidad de recopilar esta información también puede depender del CDU, ya que algunas sustancias de prueba (por ejemplo, muestras ambientales que representan mezclas complejas de composición desconocida o variable), podrían no estar bien caracterizadas.

3.3.3.2 Desarrollo de Métodos

Debe elaborarse una descripción específica, detallada y por escrito, del método con base en los datos producidos a partir de ese método. Esto puede ser en forma de un protocolo, plan de estudio, informe, y/o PNO. Cada paso del método debe investigarse para determinar en qué medida las variables ambientales, matriciales o de procedimiento, podrían afectar la detección y/o cuantificación de los analitos.

Durante el desarrollo se debe prestar mucha atención a factores como:

- Selección de reactivos (con relevancia biológica, especificidad y estabilidad apropiadas).
- Método de detección o instrumentación (es decir, procedimientos de desempeño y calibración disponibles con facilidad).
- Compatibilidad de los desechables (es decir, placas de microtitulación y otros plásticos) con las mediciones del ensayo y las sustancias de prueba.
- Método de análisis/Método estadístico.
- Pasos o procesos que podrían introducir variabilidad en el ensayo.

Se deben tomar las medidas adecuadas para minimizar los efectos externos o matriciales (o al menos caracterizar dichos efectos) en el rango de aplicación del método, especialmente si los compuestos, la matriz o los equipos utilizados durante el desarrollo son diferentes de los empleados durante la caracterización técnica del método.

3.3.3.3 Mediciones de Criterios de Valoración y Parámetros

Las mediciones de cada criterio de valoración o analito deberán estar satisfactoriamente probadas y documentadas. El desarrollo del método para una MNE novedosa debe incluir la demostración de que éste puede medir con éxito todos los parámetros relevantes. Se deben capturar y registrar metadatos relevantes de cada experimento para vincular datos cuantitativos con información cualitativa, tales como condiciones experimentales o sustancias químicas de prueba y concentraciones, y para rastrear factores externos (como la fecha o el técnico que realiza el experimento) que pudieran contribuir a la variación del ensayo o a los efectos relacionados con el lote. Estos datos y metadatos deben exportarse y guardarse en un formato accesible para que puedan ser consultados durante revisiones independientes de la validación. Este tema también se trata en las secciones del Apéndice B sobre gráficas de control y hojas de verificación.

3.3.3.4 Límites de Uso

Se deben identificar y describir claramente las fortalezas y limitaciones específicas del método de prueba. Se debe listar cualquier fuente potencial de interferencia e identificar cualquier sustancia química o clase de producto químico que pudieran interferir con la prueba. La documentación también debe identificar cualquier límite conocido sobre qué materiales se pueden probar utilizando la MNE.

3.3.3.5 Criterios de Valoración bien Definidos

Los datos generados por el método de prueba deben medir o predecir adecuadamente el criterio de valoración de interés, y dicho criterio de valoración debe definirse claramente con una explicación de la relevancia biológica, tal como se describe en la Sección 3.2. Un ejemplo de esto sería una MNE que proporciona información sobre un evento clave específico en una RRA. Los datos también deben describir cualquier vinculación entre el nuevo método de prueba y un método de prueba existente, o entre el nuevo método de prueba y los efectos en la especie objetivo. Los criterios para un resultado positivo, negativo o no concluyente en la MNE, deben definirse con claridad y evaluarse respecto al tiempo para garantizar la estabilidad del sistema.

3.3.3.6 Construcción de un Modelo Estadístico

Pueden construirse modelos estadísticos empleando datos de las pruebas intralaboratorio. Estos modelos se pueden construir mediante el uso de enfoques estadísticos Bayesianos o frecuentistas. Es posible emplear histogramas para evaluar la distribución de los datos obtenidos para medidas de control en proceso y para evaluar qué tipo de distribución (por ejemplo, distribución normal) se ajusta a los datos (ver Petersen et al., 2022b para más detalles). Es útil desarrollar modelos para calcular la variabilidad acumulativa de la MNE a partir de mediciones tanto de las sustancias de prueba como de las medidas de control en el proceso, en lugar de utilizar únicamente datos sobre la variabilidad de las sustancias de prueba. Esta información se puede utilizar para construir un modelo estadístico que pueda producir una decisión (por ejemplo, si la sustancia de prueba es positiva o negativa) y la confianza estadística para tomar esa decisión. Una simple comparación del valor de la media de una evaluación de la sustancia de prueba con

un umbral, no toma en cuenta la variabilidad de los resultados de la prueba y no puede proporcionar confianza estadística para la decisión.

Una de las preocupaciones medulares al desarrollar modelos estadísticos para MNE, es cómo diferenciar entre resultados “negativos” y los “débilmente positivos”. El umbral del modelo estadístico puede ser informado por datos *in vivo* cuando los haya disponibles (Friedman et al., 2023; Karmaus et al., 2022; Pham et al., 2020). Esto puede requerir pruebas repetidas de compuestos “límite” para evaluar la reproducibilidad de la MNE; véase por ejemplo la Directriz 497 (OCDE, 2021a). Por ejemplo, un modelo estadístico que evalúa una relación dosis-respuesta, puede usarse para evaluar el punto de partida o la concentración que causa un efecto definido (por ejemplo, el valor de la CE_{50}) y los intervalos de confianza asociados basados en los datos, alrededor de esos valores. También puede ser importante evaluar la calidad de una MNE utilizando enfoques estadísticos tales como una prueba T, el factor Z, u otros criterios estadísticos apropiados (Zhang et al., 1999; Zhang, 2011).

3.3.3.7 Reproducibilidad de los Resultados del Ensayo

Se deberá incluir documentación de reproducibilidad técnica con la información presentada, cuando sea aplicable. La reproducibilidad del método se puede evaluar mediante mediciones replicadas, incluyendo controles de calidad y muestras. Esta evaluación debe incluir la discusión de los fundamentos para la selección de las sustancias utilizadas para evaluar la reproducibilidad (posiblemente en cualquier estudio intra- e interlaboratorios realizado) y la medida en que representan el rango de posibles resultados de la prueba. Se deberán identificar y discutir los valores atípicos. Debe incluirse un análisis estadístico cuantitativo de la extensión de cualquier variabilidad intra- y/o interlaboratorios o un análisis del coeficiente de variación. Las medidas de tendencia central y variabilidad deben resumirse en los datos de control histórico (negativos, positivos y vehiculares en su caso). Cuando se prueba el (los) mismo(s) compuesto(s) varias veces, las comparaciones pueden ser cuantitativas (por ejemplo, los valores CE_{50} obtenidos) o cualitativas (por ejemplo, la clasificación de peligro). En casos en los que el método de prueba propuesto sea mecánica y funcionalmente similar a un método de prueba establecido, con estándares de desempeño existentes (por ejemplo, de una GP de la OCDE), se debe comparar la confiabilidad de los dos métodos de prueba y discutir el impacto potencial de cualquier diferencia.

3.3.3.8 Procedimiento de Interpretación de datos

El procedimiento de interpretación de datos, incluidos los criterios para respuestas positivas y negativas, debe describirse claramente para cada MNE. La combinación de MNE en EDs requiere procedimientos fijos de interpretación de datos que sean objetivos y no incluyan un juicio pericial, los cual asegura que producirán el mismo resultado cuando sean aplicados por diferentes grupos (OCDE, 2017). El uso de algoritmos computacionales, por ejemplo, modelos de aprendizaje automático y software (incluyendo el número de versión), debe estar bien documentado para garantizar la reproducibilidad de las conclusiones.

3.4 Integridad de los datos

La integridad de los datos es un aspecto clave para garantizar que la información derivada de las MNE sea creíble y confiable. Se aconseja a los desarrolladores de métodos realizar una evaluación interna de los procesos empleados para la adquisición, transferencia y procesamiento de datos sin procesar, antes de que dichos datos sean presentados a partes externas e independientes para su evaluación y revisión por colegas, para garantizar la integridad de los datos y la credibilidad de los resultados. Los estudios deben realizarse en la medida de lo posible de acuerdo con los principios de las BPL (21 CFR § 58; 40 CFR § 160; 40 CFR § 792; OCDE, 1998), cuando sea necesario. Por otra parte, los organismos evaluadores, como el Programa Nacional de Toxicología del Centro Interinstitucional para la Evaluación de Métodos Toxicológicos Alternativos (NICEATM, según sus siglas en inglés) pueden facilitar las evaluaciones de calidad e integridad del proceso de desarrollo de las MNE (NIEHS, 2023b). Hay otros recursos disponibles que proporcionan orientación sobre la maximización de la integridad de datos. Para métodos *in vitro*, los desarrolladores pueden consultar el documento GIVMP de la OCDE (OECD, 2018). Para las herramientas digitales y la gestión de datos digitales, los desarrolladores pueden seguir los “Principios de Guía FAIR para la gestión y administración de datos científicos”, publicados en 2016 (Wilkinson et al., 2016).

3.5 Transparencia de la Información

La transparencia facilita la confianza en el uso de las MNE y por lo tanto acelera el ritmo del proceso de toma de decisiones regulatorias de una institución y la potencial aceptación o calificación regulatoria. La relevancia de una MNE para la especie, el CDU y la caracterización técnica, debe comunicarse de manera transparente a colegas revisores, a la comunidad científica y al público. Cuando corresponda, los artículos revisados por colegas y la información que describa el CDU, la relevancia biológica y la caracterización técnica de la MNE, deben ser publicados en revistas de acceso abierto y/o resumidos en documentos regulatorios públicos. De manera ideal, los principios de la MNE, el protocolo, los archivos de datos sin procesar y los textos utilizados para analizar y graficar datos, además de las normas de elaboración de informes, deben comunicarse al público. Para MNE que contienen propiedad intelectual, la OCDE proporciona herramientas para mantener la transparencia, incluyendo términos razonables y no discriminatorios para los compromisos de otorgamiento de licencias (OCDE, 2021b). El uso de técnicas o equipos propietarios o patentados en un método, potencialmente puede realizarse para satisfacer las necesidades de pruebas reglamentarias. Para algunas agencias puede ser necesario que el desarrollador de la MNE transmita información propietaria o patentada para respaldar la aceptación o calificación regulatoria; una guía específica para una institución puede orientar a los desarrolladores de métodos de prueba acerca de las necesidades de información de dicha institución.

Los socios de la Cooperación Internacional sobre Métodos de Prueba Alternativos (ICATM; NIEHS, 2023c) publican información sobre la evaluación de MNE y su revisión por colegas, a través del Sistema de Seguimiento de Métodos Alternativos con miras a la Aceptación Regulatoria (TSAR, según sus siglas en inglés) (EURL ECVAM, 2021). El TSAR indica las etapas que las MNE han alcanzado en términos de aceptación como estándar reconocido para su

uso en un contexto regulatorio, junto con una descripción resumida y protocolo(s) o PNO(s) aceptados. En caso de estar disponibles, el TSAR también incluye registros y documentos relevantes asociados a una MNE vinculados a las diferentes etapas de todo el proceso: presentación, validación, revisión por colegas, recomendaciones y aceptación o calificación regulatoria. El modo de interpretar los datos que genera una MNE y los criterios de aceptación asociados, deben comunicarse claramente para que los usuarios finales entiendan el proceso y puedan aplicarlo en un entorno práctico.

3.6 Revisión Independiente

La información y los datos que sustentan el CDU de la MNE, la relevancia biológica y la caracterización técnica, pueden ser revisados científicamente por terceros independientes (cuyos miembros no tengan conflictos de intereses); sin embargo, el nivel necesario de revisión dependerá de la normativa y política de cada institución.

La evaluación de una MNE se puede lograr a través de varios procesos diferentes e implica documentar las características de desempeño de un método para determinar si son adecuados y confiables para la(s) aplicación(es) prevista(s). La confiabilidad de un método con frecuencia incluye (pero no se limita a) la reproducibilidad, repetibilidad y robustez. No obstante, puede haber información adicional no presentada en este informe, que pudiera ser necesaria para la revisión de algunos métodos. Además del desempeño y aplicabilidad de la MNE, las buenas prácticas científicas, técnicas y de calidad aseguran que el proceso de revisión independiente sea eficiente y eficaz y conduzca a una mayor confianza en el método propuesto. Los laboratorios deben conservar toda la información relevante para la evaluación de una MNE, tal como la información necesaria para operar y mantener el equipo utilizado en la ejecución de una MNE (por ejemplo, manuales de equipo y software y certificados de conformidad de calidad y seguridad), así como documentación de los proveedores de materiales, células y reactivos. Para estudios destinados a cumplir con las BPL, también puede ser relevante un informe de IQ/OQ/PQ de cada instrumento utilizado. Para estudios no-BPL, se debe conservar la documentación de la instalación y pruebas adecuadas, para demostrar que el equipo funciona según lo previsto.

Los datos sin procesar y la información que describe a la MNE deben ser accesibles para su revisión por terceros independientes y/o por los responsables de la toma de decisiones de organismos reguladores. La evaluación y revisión de las MNE por colegas independientes, puede ser organizada por instituciones de validación, tales como el NICEATM, el Laboratorio de Referencia de la Unión Europea para Alternativas a Pruebas en Animales (EURL ECVAM, según sus siglas en inglés) y su Comité de Asesoría Científica, y el Centro Japonés para la Validación de Métodos Alternativos (JaCVAM, según sus siglas en inglés). Otras instituciones u organizaciones internacionales que pueden revisar de manera independiente las MNE, incluyen el Panel de Asesoría Científica de la Ley Federal de Insecticidas, Fungicidas y Rodenticidas de los Estados Unidos, el Comité Científico de la Autoridad Europea de Seguridad Alimentaria y la OCDE. Alternativamente, el desarrollador puede financiar (pero no dirigir directamente la administración de) una revisión independiente del método. Las publicaciones revisadas por colegas son útiles para compartir información de ensayos con la comunidad científica y pueden

complementar una revisión más formal por parte de terceros independientes para respaldar la aceptación y uso del método en un contexto regulatorio.

El alcance de la revisión independiente variará según el CDU, el marco regulatorio y el método específico que se está evaluando. Parte de la información enviada por el desarrollador para su revisión independiente incluiría registros de cualquier estudio intra- o interlaboratorios, incluyendo si la MNE fue comparada o no con otra MNE, con datos *in vivo* en animales o datos en seres humanos. Estas evaluaciones pueden ayudar potencialmente a la transferibilidad interlaboratorios de la MNE.

4.0 Aceptación de MNE por Agencias Federales de los Estados Unidos

4.1 Comprensión de las Necesidades Regulatorias y Contextos de Decisión

Las agencias federales tienen diferentes autoridades para solicitar, obtener y usar datos toxicológicos. Estas difieren según los estatutos, regulaciones y la clasificación del producto. Una MNE puede ser útil y adecuada en algunos contextos estatutarios o reglamentarios, pero no en otros. En consecuencia, los desarrolladores y usuarios de una MNE deberán considerar el contexto en el que será utilizada. Algunos contextos potenciales incluirían, pero no se limitan a, EDs, IATAs o métodos independientes.

Si una MNE será aceptable para un propósito regulatorio o de otro tipo, depende de la naturaleza de la decisión que se tome, la idoneidad de la MNE para su uso previsto y la medida en que la presentación reglamentaria depende de los resultados de la MNE para respaldar la seguridad, eficacia y/o la(s) determinación(es) de riesgo. Un NAM con alta sensibilidad y baja especificidad que identifique una señal de preocupación puede ser útil para aplicaciones en las que se cribará una gran cantidad de compuestos para priorizar evaluaciones adicionales. Tal NAM puede ser de utilidad limitada al decidir qué dosis de un solo compuesto es segura para su exposición. En algunos casos, las instituciones podrían buscar señales de preocupación en situaciones de escasez de datos (identificación de peligros). En otros casos, las instituciones pueden necesitar tomar decisiones más cuantitativas sobre los compuestos (evaluación de riesgos) y podrían no estar dispuestas a aceptar un alto nivel de incertidumbre en los resultados provenientes de las MNE que se utilizan para informar tales decisiones. La implementación de las características de desempeño debe reflejar cómo la NAM será empleada e individualizada, para cada MNE y su CDU.

Para facilitar este proceso, es posible que algunas instituciones deseen evaluar las MNE antes de su uso en un proceso de toma de decisiones regulatorias. Es probable que la evaluación se centre en que la MNE tenga un CDU bien definido. Uno de los propósitos de la evaluación es hacer posible que los reguladores apliquen los resultados generados usando la MNE, sin necesidad de volver a revisar todos los datos de respaldo subyacentes de dicha MNE (véanse los documentos citados en la Tabla 1).

4.2 Consideraciones sobre el Contexto de Uso

El propósito de la MNE debe comunicarse claramente (por ejemplo, identificación de peligros, evaluación de potencia, punto de partida para evaluación cuantitativa del riesgo, etc.), y la MNE debe evaluarse en función de ese propósito.⁶ Conviene centrar la aplicación de una MNE recientemente desarrollada en un único CDU. Sin embargo, se pueden agregar CDU adicionales en momentos posteriores con los datos de respaldo adecuados. Por lo general, un CDU debe centrarse en una necesidad regulatoria particular. Los esfuerzos iniciales para calificar a una MNE pueden ser más exitosos si el CDU es estrecho. Un CDU podría ampliarse con datos adicionales, según sea apropiado. Las necesidades regulatorias difieren según las instituciones, por lo que un CDU para una MNE en particular también podría diferir. Es esencial establecer un CDU adecuado para una MNE propuesta antes de llevar a cabo un proceso completo de calificación.

Por lo general, la determinación de un CDU apropiado debe ser discutida entre el desarrollador de la MNE y la(s) institución(es) para la(s) cual(es) el CDU es relevante. Pueden ser necesarias varias iteraciones de un CDU antes de definir uno aceptable. Un CDU puede incluso cambiar durante la recopilación de datos, a medida que se definen con mayor detalle la aplicabilidad y las limitaciones de una MNE.

4.3 Evolución de la Confianza Basada en la Experiencia Adquirida

La incorporación de una MNE al uso regulatorio requiere una confianza suficiente en el método, por parte de los reguladores y de la industria regulada. La validación y la calificación respaldan esta confianza, pero podrían no ser suficientes para garantizar la implementación. Por lo general, es necesaria una educación sobre y experiencia con una MNE, antes de obtener una aceptación generalizada para su propósito. El costo, la complejidad, disponibilidad de reactivos y personal capacitado en la ejecución e interpretación de una MNE, son algunos factores que pueden limitar la adopción de la misma, incluso si se dispone de datos de validación y calificación.

Por lo general, será necesario comparar el desempeño de una MNE destinada a complementar o reemplazar un enfoque existente, con dicho enfoque existente. A menudo existe una gran confianza en los enfoques existentes con los cuales se tiene una sustancial experiencia. Es posible que dichos enfoques existentes no hayan sido objeto de validación formal, pero el uso exitoso repetido del enfoque existente junto con la supuesta validez inherente de las pruebas en animales, a menudo generan una confianza sustancial en el mencionado enfoque. Los usuarios necesitan saber que una MNE será tan buena o mejor que los enfoques existentes, cuando los resultados se emplean para tomar decisiones sobre seguridad. Pueden ser aceptables diferentes niveles de incertidumbre para diferentes CDU, en donde es necesaria una mayor confianza a medida que se avanza a través de la priorización y selección, hasta la caracterización de peligros y evaluación de riesgos, por ejemplo.

Un mecanismo para generar confianza en una MNE es que los usuarios proporcionen datos de la MNE en paralelo con datos del método existente que la MNE pretende reemplazar o complementar. El intercambio voluntario de información sobre las MNE dentro y entre

industrias, puede ayudar a establecer un volumen de datos suficientemente grande para respaldar la confianza en los métodos. Con el tiempo, las instituciones y la industria pueden darse cuenta cómo la MNE puede encajar en los paradigmas de evaluación existentes, sin comprometer las normas de seguridad.

5.0 Estados Unidos y la armonización internacional

La coordinación entre las instituciones federales de los Estados Unidos y más ampliamente con las autoridades reguladoras internacionales, ayudará a garantizar la armonización de los enfoques para validar las MNE y respaldar su aplicación e implementación. El ICCVAM y NICEATM desempeñan un papel importante en la facilitación de la comunicación y colaboración, tanto a nivel nacional como global.

5.1 Armonización de Estados Unidos: papel del ICCVAM y del NICEATM

La Ley de Autorización del ICCVAM señala los siguientes propósitos del mismo (42 U.S.C 2851-3, 2000; NIEHS, 2023d):

- Aumentar la eficiencia y efectividad de la revisión de métodos de prueba de la agencia federal de los Estados Unidos.
- Eliminar la duplicación innecesaria de esfuerzos y compartir experiencias entre las instituciones reguladoras federales de los Estados Unidos.
 - Esto se logra a través de diversos medios, como reuniones mensuales, grupos de trabajo, reuniones abiertas al público (Foro Público) y grupos de asesoría científica (Comité de Asesoría Científica sobre Métodos Toxicológicos Alternativos).
- Optimizar el uso de la experiencia científica fuera del Gobierno Federal de los Estados Unidos.
 - Esto con frecuencia se logra mediante la organización de reuniones y talleres con asociados de la ICATM, así como con numerosos científicos en conferencias.
- Asegurar que los métodos de prueba nuevos y revisados sean validados para satisfacer las necesidades de las agencias federales de los Estados Unidos.
 - Los científicos y reguladores de agencias federales cooperan con los desarrolladores para garantizar que los métodos que se están produciendo atiendan una necesidad regulatoria.
- Reducir, refinar o reemplazar el uso de animales en pruebas, cuando sea posible.

El ICCVAM facilita colaboraciones interinstitucionales e internacionales promoviendo el desarrollo, aceptación o calificación regulatoria y uso de pruebas alternativas que fomenten la reducción, refinamiento o sustitución de métodos de prueba en animales. El ICCVAM brinda orientación a los desarrolladores de métodos de prueba, evalúa recomendaciones de revisiones realizadas por colegas expertos sobre métodos alternativos de pruebas toxicológicas y hace recomendaciones sobre el uso de métodos de prueba revisados, a las instituciones federales

apropiadas. El ICCVAM desempeña sus funciones a través de grupos de trabajo técnicos ad hoc, gestionados por el NICEATM, para realizar tareas específicas importantes para el desarrollo o validación de alternativas a las pruebas en animales. Un ejemplo en curso es el apoyo del ICCVAM en la coordinación de un estudio de prevalidación interlaboratorios, para una MNE desarrollada por la EPA basada en un ensayo *in vitro* de microtejido tiroideo humano para detección química (Deisenroth et al., 2020).

El NICEATM, que es una oficina de la División de Toxicología Traslacional del Instituto Nacional de Ciencias de la Salud Ambiental (NIEHS, según sus siglas en inglés), brinda apoyo técnico, científico y operativo al ICCVAM y a las actividades de grupos de trabajo del ICCVAM, paneles de revisión por colegas, paneles de expertos, talleres y esfuerzos de validación. Además de apoyar al ICCVAM, el NICEATM:

- Realiza análisis y evaluaciones de métodos de prueba y coordina estudios de validación independientes, sobre enfoques de prueba alternativos novedosos y de alta prioridad.
- Proporciona información a desarrolladores de métodos de prueba, reguladores y la industria regulada, a través del sitio web del NICEATM, el Entorno Químico Integrado y talleres sobre temas de interés.
- Apoya actividades de la División de Toxicología Traslacional del NIEHS, especialmente aquellas que contribuyen con el consorcio Interinstitucional de Toxicología en el Siglo XXI (Tox21) del gobierno de los Estados Unidos.

El foro de comunicación interinstitucional proporcionado por el ICCVAM y el apoyo brindado por NICEATM, sirven para garantizar que los recursos limitados se estén aprovechando de manera efectiva para coordinar los esfuerzos de las instituciones federales de los Estados Unidos para validar y calificar MNE para aplicación regulatoria.

5.2 Armonización de Estados Unidos: Colaboraciones Federales Adicionales para Avanzar las 3Rs

Las agencias federales de los Estados Unidos colaboran en las MNE de muchas maneras, además de su participación en el ICCVAM. Por ejemplo, Tox21 (Tox21, n.d.) es una colaboración federal entre la EPA, la División de Toxicología Traslacional del NIEHS, el Centro Nacional para el Avance de las Ciencias Traslacionales (NCATS) dentro de los Institutos Nacionales de Salud, y la FDA. El objetivo de Tox21 es desarrollar mejores métodos de evaluación de toxicidad, para probar de manera eficiente si ciertos compuestos químicos pueden tener el potencial de interrumpir procesos biológicos en el cuerpo humano y conducir a efectos negativos para la salud. Tox21 ha producido una serie de publicaciones y análisis seminales que se han puesto en uso regulatorio, por ejemplo en el Programa de Detección de Disruptores Endocrinos de la EPA (EPA, 2023). El NICEATM y la EPA también han trabajado estrechamente en múltiples estudios para reducir o reemplazar el uso de animales en pruebas regulatorias. Esto incluye análisis retrospectivos para eliminar el uso de animales para: 1) toxicidad aguda dérmica de plaguicidas y formulaciones plaguicidas (EPA, 2020); 2) determinar si el uso de datos *in vitro* por sí solos será suficiente para la derivación del factor de absorción dérmica, para evaluación de riesgos para la salud humana de los plaguicidas (Allen et al., 2021); y 3) determinar si se puede

lograr el mismo nivel de protección de vertebrados acuáticos no-objetivo, con pruebas de toxicidad aguda *in vivo* en menos de tres especies de peces (Ceger et al., 2023). La FDA también colabora con el NCATS para desarrollar aún más tecnologías de sistemas microfisiológicos, para promover su avance y acelerar el uso traslacional (FDA, 2023).

5.3 Armonización Internacional

Además de los esfuerzos de colaboración internacional facilitados a través de ICCVAM, las instituciones estadounidenses colaboran de manera independiente a nivel internacional para avanzar en la aceptación de las MNE. Algunos ejemplos incluyen el compromiso del ICATM de los Estados Unidos con el subcomité de expertos de las Naciones Unidas sobre el Sistema Globalmente Armonizado (GHS, según sus siglas en inglés) de Clasificación y Etiquetado de Sustancias Químicas, la colaboración con el ICH para elaborar directrices que describan el uso de las MNE y la participación en el Programa de Directrices de Prueba de la OCDE y en el Grupo de Trabajo sobre Evaluación de Peligros.

El ICATM fue establecido como una asociación entre organizaciones de validación de los Estados Unidos (ICCVAM), Japón (JaCVAM), la Unión Europea (EURL ECVAM) y Canadá (Ciencias de la Salud Ambiental y Oficina de Investigación, dentro de Salud Canadá). Otras organizaciones participantes incluyen el Centro Coreano para la Validación de Métodos Alternativos, el Centro Brasileño para la Validación de Métodos Alternativos, y la Administración China de Alimentos y Medicamentos, y el Centro de Control y Prevención de Enfermedades de Guangdong. Los objetivos generales de este grupo han sido:

- Establecer la cooperación internacional en áreas críticas de estudios de validación, la revisión independiente por colegas y el desarrollo de recomendaciones armonizadas, para asegurar que los métodos/estrategias alternativos sean más fácilmente aceptados en todo el mundo.
- Establecer la cooperación internacional necesaria para garantizar que los nuevos métodos/estrategias de prueba alternativos adoptados para uso regulatorio, brinden una protección equivalente o mejorada a personas, animales y el medio ambiente, a la vez que reemplazan, reducen o refinan (causando menos dolor y angustia) el uso de animales, siempre que sea científicamente factible.

Dentro de las Naciones Unidas, el subcomité del GHS ha establecido un grupo de trabajo para actualizar varios capítulos del GHS, a fin de establecer criterios específicos para el uso de MNE en diversas clases de peligros (por ejemplo, irritación/corrosión en la piel, irritación ocular/daño ocular grave, sensibilización cutánea). Estos esfuerzos han sido fructíferos en el avance de las MNE a nivel internacional y han establecido EDs como métodos aceptables para la determinación de peligros. Estos esfuerzos internacionales son importantes no sólo para promover las MNE, sino también para proporcionar conocimientos especializados a regiones internacionales que tal vez no dispongan de recursos suficientes en este campo.

El valor de armonizar los enfoques al uso de MNE también se ilustra en las actividades del ICH. Varias guías desarrolladas por el ICH describen el uso de métodos alternativos que son

considerados enfoques aceptables por múltiples autoridades reguladoras y grupos de la industria en todo el mundo. Estos métodos fueron evaluados por grupos de trabajo de expertos dentro del proceso del ICH e incorporados a las directrices, según corresponda. Algunos ejemplos incluyen el uso de métodos *in quimico* e *in vitro* para la evaluación de fototoxicidad (ICH, 2013).

Además, la guía del ICH sobre toxicidad reproductiva y del desarrollo incluye algunos contextos de uso de ensayos alternativos, así como recomendaciones sobre el enfoque para calificar tales ensayos y una lista de compuestos de referencia de criterio de valoración biológico (FDA, 2021b).

La OCDE es un foro internacional para armonizar las directrices de pruebas regulatorias y los documentos de orientación, y se centra cada vez más en la validación y el uso de MNE. El ICCVAM desempeña un papel importante en la coordinación y contribución a la postura de Estados Unidos sobre el Programa de Guías de Pruebas de Efectos en la Salud de la OCDE. El Coordinador Nacional de los Estados Unidos representa a dicho país en la reunión anual del Grupo de Trabajo de Coordinadores Nacionales y en otras actividades de desarrollo de directrices de pruebas. En papel, el Coordinador Nacional de Estados Unidos solicita aportes de instituciones pertinentes del ICCVAM, para las actividades de GP de la OCDE que involucren cualquier aspecto de las 3Rs. Expertos en la materia de las agencias del ICCVAM prestan servicio en múltiples grupos de expertos de la OCDE, para proporcionar orientación científica sobre el desarrollo de productos de la OCDE tales como directrices de prueba, EDs y documentos de guía. Las directrices de prueba de la OCDE son utilizadas por las partes interesadas de los 38 países miembros de la OCDE, para evaluar la seguridad química. La cláusula de Aceptación Mutua de Datos de la OCDE garantiza que los datos de seguridad generados mediante una guía de prueba de la OCDE, serán aceptados por todos los países miembro, lo cual evita pruebas redundantes. Las agencias del ICCVAM también contribuyen al Proyecto de Estudios de Casos de IATA de la OCDE, que permite a los países compartir y colaborar en el uso de metodologías novedosas en los IATA para evaluar la seguridad química dentro de un contexto regulatorio.

6.0 Comunicación y Capacitación para Fomentar el Uso de MNE

La comunicación de las instituciones sobre la aceptabilidad de MNE específicas y la capacitación sobre las MNE pueden facilitar su uso. Cuando sea apropiado y factible, las instituciones podrían comunicar públicamente cuándo y cómo una MNE es aceptable, dependiendo de las normas y políticas específicas de cada institución. Por ejemplo, las MNE pueden describirse en DG o en sitios web disponibles al público. Las instituciones reguladoras pueden usar programas de capacitación existentes o implementar nuevos programas para el personal y proporcionar capacitación disponible al público, cuando sea posible, sobre los nuevos métodos. La creación de confianza en nuevos enfoques puede comenzar incluso antes de que una MNE sea validada o calificada, mediante la educación de la comunidad científica. Estos esfuerzos de educación temprana pueden centrarse en la ciencia básica de los nuevos enfoques. A medida que los enfoques maduran y los datos que respaldan la validez de un enfoque se acumulan, los esfuerzos educativos pueden familiarizar a la comunidad sobre estos datos. La

capacitación sobre el uso específico e interpretación de MNE puede llevarse a cabo cuando las MNE han sido evaluadas para CDU particulares. Entidades como la OCDE, el ICCVAM y otras organizaciones de validación y sociedades científicas, también pueden brindar capacitación y acceso a información sobre MNE, lo cual respaldará la confianza en su uso.

La interacción entre desarrolladores de MNE, usuarios de la industria y reguladores puede facilitar el desarrollo y la adopción de las metodologías. Dichas interacciones pueden ocurrir a través de la participación de todas las partes en reuniones científicas, en donde se discutan dichos métodos y a través de interacciones regulatorias más formales de acuerdo con los procesos específicos de la agencia. Como se señaló anteriormente, el desarrollo de CDU apropiados y conjuntos de datos de calificación puede ser un proceso iterativo. Los CDU y los dominios de aplicabilidad pueden cambiar durante el desarrollo y exploración de una MNE a medida que se acumulan los datos. La comunicación continua entre todas las partes durante este proceso, puede ayudar a asegurar que se tome un camino adecuado para la aceptación de una MNE.

Incluso si se dispone de una MNE para un criterio de valoración y es aceptada por una institución reguladora, el patrocinador de una solicitud puede optar por utilizar otros enfoques, tales como una prueba tradicional en animales, por lo que la MNE no siempre puede presentarse en una solicitud reglamentaria. Podría haber varias razones para ello que están fuera del control de las autoridades reguladoras. Aunque una agencia reguladora puede sugerir o recomendar la sustitución de una prueba tradicional por una o varias MNE, el patrocinador de un compuesto podría no estar siempre obligado a seguir la sugerencia. La educación y familiaridad con las MNE entre todas las partes interesadas, son necesarias para generar suficiente confianza para la adopción de las MNE.

7.0 Conclusión e implementación

El presente informe está destinado a ayudar a los desarrolladores de métodos, a las partes interesadas de las industrias reguladas y a las instituciones federales, en el desarrollo, validación, calificación y aceptación de MNE científicamente relevantes. Aquí, hemos descrito conceptos clave que se deben tomar en cuenta para permitir el desarrollo eficiente y oportuno de MNE aptas para su propósito, confiables y que proporcionen información relevante para las especies de interés. Toda la información puede aplicarse o no a cualquier método específico, ED o IATA. También puede haber otros conceptos que se apliquen a un método, ED o IATA, que no fueron discutidos en este informe. Es importante que los desarrolladores trabajen en estrecha colaboración con las instituciones federales y los usuarios finales, con la vista puesta en el uso previsto de la MNE, particularmente para aplicaciones de evaluación de riesgos en una revisión regulatoria. Establecer la confianza científica en las MNE y validar o calificar métodos para fines y CDU específicos, deben ser procesos iterativos que evolucionen a través de una comunicación multidireccional entre las partes interesadas.

El campo de las MNE está en evolución y pueden surgir nuevas consideraciones sobre la validación y calificación de MNE que no se habían anticipado en el momento de la redacción de

este informe. En consecuencia, es posible que las partes interesadas que participan en el desarrollo, validación y calificación de las MNE, deban permanecer flexibles y abiertas a incorporar consideraciones no descritas en el presente documento. Este informe se actualizará periódicamente y según sea necesario.

Referencias

- 15 USC §2601, 2016. 15 USC §2601: Findings, policy, and intent [WWW Document]. URL [https://uscode.house.gov/view.xhtml?req=\(title:15%20section:2601%20edition:prelim\)](https://uscode.house.gov/view.xhtml?req=(title:15%20section:2601%20edition:prelim)) (accessed 11.2.22).
- 21 CFR § 58, n.d. CFR - Code of Federal Regulations Title 21 [WWW Document]. URL <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=58> (accessed 11.3.22).
- 40 CFR § 160, n.d. CFR - Code of Federal Regulations Title 40 [WWW Document]. URL <https://www.ecfr.gov/current/title-40/chapter-I/subchapter-E/part-160> (accessed 5.9.23).
- 40 CFR § 792, n.d. CFR - Code of Federal Regulations Title 40 [WWW Document]. URL <https://www.ecfr.gov/current/title-40/chapter-I/subchapter-R/part-792> (accessed 5.9.23).
- 42 U.S.C 2851-3, 2000. ICCVAM Authorization Act of 2000 [WWW Document]. 42 U.S.C. 2851-3 Public Law 106-545. URL <https://www.congress.gov/bill/106th-congress/house-bill/4281/text> (accessed 5.9.23).
- Allen, D.G., Rooney, J., Kleinstreuer, N., Lowit, A., Perron, M., 2021. Retrospective analysis of dermal absorption triple pack data. *ALTEX - Alternatives to animal experimentation* 38, 463–476. <https://doi.org/10.14573/altex.2101121>
- Browne, P., Delrue, N., Gourmelon, A., 2019. Regulatory use and acceptance of alternative methods for chemical hazard identification. *Current Opinion in Toxicology* 15, 18–25.
- Browne, P., Kleinstreuer, N.C., Ceger, P., Deisenroth, C., Baker, N., Markey, K., Thomas, R.S., Judson, R.J., Casey, W., 2018. Development of a curated Hershberger database. *Reproductive Toxicology* 81, 259–271. <https://doi.org/10.1016/j.reprotox.2018.08.016>
- Ceger, P., Allen, D., Blankinship, A., Choksi, N., Daniel, A., Eckel, W.P., Hamm, J., Harwood, D.E., Johnson, T., Kleinstreuer, N., Sprankle, C.S., Truax, J., Lowit, M., 2023. Evaluation of the fish acute toxicity test for pesticide registration. *Regulatory Toxicology and Pharmacology* 139, 105340. <https://doi.org/10.1016/j.yrtph.2023.105340>
- Ceger, P., Garcia-Reyero Vinas, N., Allen, D., Arnold, E., Bloom, R., Brennan, J.C., Clarke, C., Eisenreich, K., Fay, K., Hamm, J., Henry, P.F.P., Horak, K., Hunter, W., Judkins, D., Klein, P., Kleinstreuer, N., Koehn, K., LaLone, C.A., Laurenson, J.P., Leet, J.K., Lowit,

- A., Lynn, S.G., Norberg-King, T., Perkins, E.J., Petersen, E.J., Rattner, B.A., Sprankle, C.S., Steeger, T., Warren, J.E., Winfield, S., Odenkirchen, E., 2022. Current ecotoxicity testing needs among selected U.S. federal agencies. *Regulatory Toxicology and Pharmacology* 133, 105195. <https://doi.org/10.1016/j.yrtph.2022.105195>
- Chang, X., Tan, Y.-M., Allen, D.G., Bell, S., Brown, P.C., Browning, L., Ceger, P., Gearhart, J., Hakkinen, P.J., Kabadi, S.V., Kleinstreuer, N.C., Lumen, A., Matheson, J., Paini, A., Pangburn, H.A., Petersen, E.J., Reinke, E.N., Ribeiro, A.J.S., Sipes, N., Sweeney, L.M., Wambaugh, J.F., Wange, R., Wetmore, B.A., Mumtaz, M., 2022. IVIVE: Facilitating the use of in vitro toxicity data in risk assessment and decision making. *Toxics* 10, 232. <https://doi.org/10.3390/toxics10050232>
- Chiu, W.A., Wright, F.A., Rusyn, I., 2017. A tiered, Bayesian approach to estimating of population variability for regulatory decision-making. *ALTEX* 34, 377–388. <https://doi.org/10.14573/altex.1608251>
- Choksi, N.Y., Truax, J., Layton, A., Matheson, J., Mattie, D., Varney, T., Tao, J., Yozzo, K., McDougal, A.J., Merrill, J., Lowther, D., Barroso, J., Linke, B., Casey, W., Allen, D., 2019. United States regulatory requirements for skin and eye irritation testing. *Cutan Ocul Toxicol* 38, 141–155. <https://doi.org/10.1080/15569527.2018.1540494>
- Church, R.J., Gatti, D.M., Urban, T.J., Long, N., Yang, X., Shi, Q., Eaddy, J.S., Mosedale, M., Ballard, S., Churchill, G.A., Navarro, V., Watkins, P.B., Threadgill, D.W., Harrill, A.H., 2015. Sensitivity to hepatotoxicity due to epigallocatechin gallate is affected by genetic background in diversity outbred mice. *Food Chem Toxicol* 76, 19–26. <https://doi.org/10.1016/j.fct.2014.11.008>
- Clippinger, A.J., Raabe, H.A., Allen, D.G., Choksi, N.Y., van der Zalm, A.J., Kleinstreuer, N.C., Barroso, J., Lowit, A.B., 2021. Human-relevant approaches to assess eye corrosion/irritation potential of agrochemical formulations. *Cutan Ocul Toxicol* 40, 145–167. <https://doi.org/10.1080/15569527.2021.1910291>
- Corley, R.A., Kuprat, A.P., Suffield, S.R., Kabilan, S., Hinderliter, P.M., Yugulis, K., Ramanarayanan, T.S., 2021. New approach methodology for assessing inhalation risks of a contact respiratory cytotoxicant: computational fluid dynamics-based aerosol dosimetry modeling for cross-species and in vitro comparisons. *Toxicological Sciences* 182, 243–259. <https://doi.org/10.1093/toxsci/kfab062>
- CPSC, 2020. Proposed Guidance for Industry and Test Method Developers: CPSC Staff Evaluation of Alternative Test Methods and Integrated Testing Approaches and Data Generated from Such Methods to Support FHSA Labeling Requirements [WWW Document]. URL <https://www.regulations.gov/document/CPSC-2021-0006-0001> (accessed 5.4.22).
- CPSC, 2012. Recommended Procedures Regarding the CPSC’s Policy on Animal Testing [WWW Document]. URL <https://www.cpsc.gov/Business--Manufacturing/Testing->

- Certification/Recommended-Procedures-Regarding-the-CPSCs-Policy-on-Animal-Testing (accessed 5.4.22).
- Crofton, K.M., Mundy, W.R., 2021. External Scientific Report on the Interpretation of Data from the Developmental Neurotoxicity In Vitro Testing Assays for Use in Integrated Approaches for Testing and Assessment. EFSA Supporting Publications 18, 6924E. <https://doi.org/10.2903/sp.efsa.2021.EN-6924>
- Daniel, A.B., Strickland, J., Allen, D., Casati, S., Zuang, V., Barroso, J., Whelan, M., Régimbald-Krnel, M.J., Kojima, H., Nishikawa, A., Park, H.-K., Lee, J.K., Kim, T.S., Delgado, I., Rios, L., Yang, Y., Wang, G., Kleinstreuer, N., 2018. International regulatory requirements for skin sensitization testing. *Regul Toxicol Pharmacol* 95, 52–65. <https://doi.org/10.1016/j.yrtph.2018.03.003>
- Deisenroth, C., Soldatow, V.Y., Ford, J., Stewart, W., Brinkman, C., LeCluyse, E.L., MacMillan, D.K., Thomas, R.S., 2020. Development of an *In Vitro* Human Thyroid Microtissue Model for Chemical Screening. *Toxicological Sciences* 174, 63–78. <https://doi.org/10.1093/toxsci/kfz238>
- Dumont, J., Ewart, D., Mei, B., Estes, S., Kshirsagar, R., 2016. Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. *Crit Rev Biotechnol* 36, 1110–1122. <https://doi.org/10.3109/07388551.2015.1084266>
- Elliott, J.T., Rösslein, M., Song, N.W., Toman, B., Ovaskainen, A.K.-, Maniratanachote, R., Salit, M.L., Petersen, E.J., Sequeira, F., Romsos, E., Kim, S.J., Lee, J., Moos, N.R. von, Rossi, F., Hirsch, C., Krug, H.F., Suchaoin, W., Wick, P., 2017. Toward achieving harmonization in a nanocytotoxicity assay measurement through an interlaboratory comparison study. *ALTEX - Alternatives to animal experimentation* 34, 201–218. <https://doi.org/10.14573/altex.1605021>
- EPA, 2023. Availability of New Approach Methodologies (NAMs) in the Endocrine Disruptor Screening Program (EDSP) [WWW Document]. URL <https://www.regulations.gov/document/EPA-HQ-OPP-2021-0756-0002> (accessed 2.20.23).
- EPA, 2021a. EPA New Approach Methods Work Plan [WWW Document]. URL https://www.epa.gov/system/files/documents/2021-11/nams-work-plan_11_15_21_508-tagged.pdf (accessed 5.5.22).
- EPA, 2021b. EPA Strategic Plan to Reduce the Use of Vertebrate Animals in Chemical Testing [WWW Document]. URL <https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/strategic-plan-reduce-use-vertebrate-animals-chemical> (accessed 5.5.22).
- EPA, 2021c. Chlorothalonil: Revised Human Health Draft Risk Assessment for Registration Review [WWW Document]. URL <https://www.regulations.gov/document/EPA-HQ-OPP-2011-0840-0080> (accessed 3.13.23).

- EPA, 2020. Guidance for Waiving Acute Dermal Toxicity Tests for Pesticide Technical Chemicals & Supporting Retrospective Analysis [WWW Document]. URL <https://www.regulations.gov/document/EPA-HQ-OPP-2016-0093-0181> (accessed 3.7.23).
- EPA, 2018. Strategic Plan to Promote the Development and Implementation of Alternative Test Methods Within the TSCA Program.
- EPA, 2015. Use of an Alternate testing framework for classification of eye irritation [WWW Document]. URL https://www.epa.gov/sites/default/files/2015-05/documents/eye_policy2015update.pdf (accessed 11.13.23).
- EURL ECVAM, 2021. EU Reference Laboratory for alternatives to animal testing (EURL ECVAM) [WWW Document]. URL https://joint-research-centre.ec.europa.eu/eu-reference-laboratory-alternatives-animal-testing-eurl-ecvam_en (accessed 11.3.22).
- Farmahin, R., Manning, G.E., Crump, D., Wu, D., Mundy, L.J., Jones, S.P., Hahn, M.E., Karchner, S.I., Giesy, J.P., Bursian, S.J., Zwiernik, M.J., Fredricks, T.B., Kennedy, S.W., 2013. Amino Acid Sequence of the Ligand-Binding Domain of the Aryl Hydrocarbon Receptor 1 Predicts Sensitivity of Wild Birds to Effects of Dioxin-Like Compounds. *Toxicological Sciences* 131, 139–152. <https://doi.org/10.1093/toxsci/kfs259>
- FDA, 2023. MOU 225-23-003 Memorandum of understanding between the National Institutes of Health (NCATS) and the Food and Drug Administration (FDA) for the Microphysiological Systems Program [WWW Document]. FDA. URL <https://www.fda.gov/about-fda/domestic-mous/mou-225-23-003> (accessed 5.9.23).
- FDA, 2021a. Advancing New Alternative Methodologies at FDA [WWW Document]. URL <https://www.fda.gov/media/144891/download> (accessed 7.19.23).
- FDA, 2021b. S5(R3) Detection of Reproductive and Developmental Toxicity for Human Pharmaceuticals Guidance for Industry [WWW Document]. URL <https://www.fda.gov/media/148475/download>
- FDA, 2020. Qualification Process for Drug Development Tools Guidance for Industry and FDA Staff [WWW Document]. URL <https://www.fda.gov/media/133511/download> (accessed 7.19.23).
- FDA, 2018. Bioanalytical Method Validation Guidance for Industry [WWW Document]. URL <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/bioanalytical-method-validation-guidance-industry> (accessed 7.31.23).
- FDA, 2017a. FDA’s Predictive Toxicology Roadmap [WWW Document]. URL <https://www.fda.gov/media/109634/download> (accessed 5.5.22).

- FDA, 2017b. Qualification of Medical Device Development Tools: Guidance for Industry, Tool Developers, and Food and Drug Administration Staff [WWW Document]. FDA. URL <https://www.fda.gov/medical-devices/science-and-research-medical-devices/medical-device-development-tools-mddt> (accessed 5.5.22).
- Frick, A., Suzuki, O.T., Benton, C., Parks, B., Fedoriw, Y., Richards, K.L., Thomas, R.S., Wiltshire, T., 2015. Identifying genes that mediate anthracycline toxicity in immune cells. *Front Pharmacol* 6, 62. <https://doi.org/10.3389/fphar.2015.00062>
- Friedman, K.P., Foster, M.J., Pham, L.L., Feshuk, M., Watford, S.M., Wambaugh, J.F., Judson, R.S., Setzer, R.W., Thomas, R.S., 2023. Reproducibility of organ-level effects in repeat dose animal studies. *Comput Toxicol* 28, 1–17. <https://doi.org/10.1016/j.comtox.2023.100287>
- Harrill, A.H., 2020. ToxPoint: In the era of precision medicine, diversity should not be neglected in chemical safety assessment. *Toxicol Sci* 173, 3–4. <https://doi.org/10.1093/toxsci/kfz232>
- Harrill, A.H., McAllister, K.A., 2017. New rodent population models may inform human health risk assessment and identification of genetic susceptibility to environmental exposures. *Environ Health Perspect* 125, 086002. <https://doi.org/10.1289/EHP1274>
- Hartung, T., 2010. Evidence-based toxicology - the toolbox of validation for the 21st century? *ALTEX* 27, 253–263. <https://doi.org/10.14573/altex.2010.4.253>
- Hoffmann, S., Kleinstreuer, N., Alépée, N., Allen, D., Api, A.M., Ashikaga, T., Clouet, E., Cluzel, M., Desprez, B., Gellatly, N., Goebel, C., Kern, P.S., Klaric, M., Kühnl, J., Lalko, J.F., Martinozzi-Teissier, S., Mewes, K., Miyazawa, M., Parakhia, R., van Vliet, E., Zang, Q., Petersohn, D., 2018. Non-animal methods to predict skin sensitization (I): the Cosmetics Europe database. *Crit Rev Toxicol* 48, 344–358. <https://doi.org/10.1080/10408444.2018.1429385>
- Hoffmann, S., Saliner, A.G., Patlewicz, G., Eskes, C., Zuang, V., Worth, A.P., 2008. A feasibility study developing an integrated testing strategy assessing skin irritation potential of chemicals. *Toxicol Lett* 180, 9–20. <https://doi.org/10.1016/j.toxlet.2008.05.004>
- ICCVAM, 2022. Symposium Webinar: Using New Approach Methodologies to Address Variability and Susceptibility Across Populations [WWW Document]. URL <https://ntp.niehs.nih.gov/go/popvar> (accessed 3.13.23).
- ICCVAM, 2018. A Strategic Roadmap for Establishing New Approaches to Evaluate the Safety of Chemicals and Medical Products in the United States [WWW Document]. URL https://ntp.niehs.nih.gov/iccvam/docs/roadmap/iccvam_strategicroadmap_january2018_documento_508.pdf (accessed 5.3.22).

- ICCVAM, 2003. ICCVAM Guidelines for the Nomination and Submission of New, Revised, and Alternative Test Methods [WWW Document]. URL https://ntp.niehs.nih.gov/iccvam/suppdocs/subguidelines/sd_subg034508.pdf (accessed 5.3.22).
- ICCVAM, 1997. Validation and Regulatory Acceptance of Toxicological Test Methods [WWW Document]. URL https://ntp.niehs.nih.gov/iccvam/docs/about_docs/validate.pdf (accessed 3.7.23).
- ICH, 2013. ICH Harmonized Tripartite Guideline Photosafety Evaluation of Pharmaceuticals S10 [WWW Document]. URL https://database.ich.org/sites/default/files/S10_Guideline.pdf
- Ishikawa, K., 1985. What is Total Quality Control? The Japanese Way. Translated by Lu, David J., 1st ed. Prentice-Hall.
- Judson, R.S., Magpantay, F.M., Chickarmane, V., Haskell, C., Tania, N., Taylor, J., Xia, M., Huang, R., Rotroff, D.M., Filer, D.L., Houck, K.A., Martin, M.T., Sipes, N., Richard, A.M., Mansouri, K., Setzer, R.W., Knudsen, T.B., Crofton, K.M., Thomas, R.S., 2015. Integrated model of chemical perturbations of a biological pathway using 18 in vitro high-throughput screening assays for the estrogen receptor. *Toxicol Sci* 148, 137–154. <https://doi.org/10.1093/toxsci/kfv168>
- Judson, R.S., Thomas, R.S., Baker, N., Simha, A., Howey, X.M., Marable, C., Kleinstreuer, N.C., Houck, K.A., 2019. Workflow for defining reference chemicals for assessing performance of in vitro assays. *ALTEX* 36, 261. <https://doi.org/10.14573/altex.1809281>
- Karmaus, A.L., Mansouri, K., To, K.T., Blake, B., Fitzpatrick, J., Strickland, J., Patlewicz, G., Allen, D., Casey, W., Kleinstreuer, N., 2022. Evaluation of variability across rat acute oral systemic toxicity studies. *Toxicological Sciences* 188, 34–47. <https://doi.org/10.1093/toxsci/kfac042>
- Kleinstreuer, N.C., Ceger, P., Watt, E.D., Martin, M., Houck, K., Browne, P., Thomas, R.S., Casey, W.M., Dix, D.J., Allen, D., Sakamuru, S., Xia, M., Huang, R., Judson, R., 2017. Development and validation of a computational model for androgen receptor activity. *Chem. Res. Toxicol.* 30, 946–964. <https://doi.org/10.1021/acs.chemrestox.6b00347>
- Kleinstreuer, N.C., Hoffmann, S., Alépée, N., Allen, D., Ashikaga, T., Casey, W., Clouet, E., Cluzel, M., Desprez, B., Gellatly, N., Göbel, C., Kern, P.S., Klaric, M., Kühnl, J., Martinozzi-Teissier, S., Mewes, K., Miyazawa, M., Strickland, J., van Vliet, E., Zang, Q., Petersohn, D., 2018. Non-animal methods to predict skin sensitization (II): an assessment of defined approaches *. *Crit Rev Toxicol* 48, 359–374. <https://doi.org/10.1080/10408444.2018.1429386>
- Kolle, S.N., Van Cott, A., van Ravenzwaay, B., Landsiedel, R., 2017. Lacking applicability of in vitro eye irritation methods to identify seriously eye irritating agrochemical formulations:

- Results of bovine cornea opacity and permeability assay, isolated chicken eye test and the EpiOcular; ET-50 method to classify according to UN GHS. *Regulatory Toxicology and Pharmacology* 85, 33–47. <https://doi.org/10.1016/j.yrtph.2017.01.013>
- Krishna, S., Berridge, B., Kleinstreuer, N., 2021. High-throughput screening to identify chemical cardiotoxic potential. *Chem Res Toxicol* 34, 566–583. <https://doi.org/10.1021/acs.chemrestox.0c00382>
- LaLone, C.A., Villeneuve, D.L., Lyons, D., Helgen, H.W., Robinson, S.L., Swintek, J.A., Saari, T.W., Ankley, G.T., 2016. Editor's Highlight: Sequence Alignment to Predict Across Species Susceptibility (SeqAPASS): A Web-Based Tool for Addressing the Challenges of Cross-Species Extrapolation of Chemical Toxicity. *Toxicol. Sci.* 153, 228–245. <https://doi.org/10.1093/toxsci/kfw119>
- Luechtefeld, T., Maertens, A., Russo, D.P., Rovida, C., Zhu, H., Hartung, T., 2016. Analysis of public oral toxicity data from REACH registrations 2008-2014. *ALTEX* 33, 111–122. <https://doi.org/10.14573/altex.1510054>
- Madia, F., Pillo, G., Worth, A., Corvi, R., Prieto, P., 2021. Integration of data across toxicity endpoints for improved safety assessment of chemicals: the example of carcinogenicity assessment. *Arch Toxicol* 95, 1971–1993. <https://doi.org/10.1007/s00204-021-03035-x>
- NIEHS, 2023a. Testing Regulations and Guidelines [WWW Document]. National Toxicology Program. URL <https://ntp.niehs.nih.gov/go/837330> (accessed 7.19.23).
- NIEHS, 2023b. Funding Opportunities for Test Method Developers [WWW Document]. Funding Opportunities for Test Method Developers. URL <https://ntp.niehs.nih.gov/go/alt-funding> (accessed 5.9.23).
- NIEHS, 2023c. International Cooperation on Alternative Test Methods [WWW Document]. International Cooperation on Alternative Test Methods. URL <https://ntp.niehs.nih.gov/go/icatm> (accessed 5.9.23).
- NIEHS, 2023d. About ICCVAM [WWW Document]. About ICCVAM. URL <https://ntp.niehs.nih.gov/go/iccvam> (accessed 5.9.23).
- OECD, 2023. No. 377: Initial Recommendations on Evaluation of Data from the Developmental Neurotoxicity (DNT) In-Vitro Testing Battery, OECD Series on Testing and Assessment. OECD Publishing, Paris.
- OECD, 2022a. No. 364: Case study on the use of Integrated Approaches for Testing and Assessment for DNT to prioritize a class of Organophosphorus flame retardants, Series on Testing and Assessment. OECD Publishing, Paris.
- OECD, 2022b. No. 367: Case Study on the use of an Integrated Approach for Testing and Assessment (IATA) for New Approach Methodology (NAM) for Refining Inhalation

- Risk Assessment from Point of Contact Toxicity of the Pesticide, Chlorothalonil, OECD Series on Testing and Assessment. OECD Publishing, Paris.
- OECD, 2021a. Guideline No. 497: Defined Approaches on Skin Sensitisation, OECD Guidelines for the Testing of Chemicals, Section 4. OECD Publishing, Paris.
- OECD, 2021b. No. 298: Guiding Principles on Good Practices for the Availability/Distribution of Protected Elements in OECD Test Guidelines, Series on Testing and Assessment. OECD Publishing.
- OECD, 2018. No. 286: Guidance Document on Good In Vitro Method Practices (GIVIMP), OECD Series on Testing and Assessment. OECD Publishing.
<https://doi.org/10.1787/9789264304796-en>
- OECD, 2017. No. 255: Guidance Document on the Reporting of Defined Approaches to be Used Within Integrated Approaches to Testing and Assessment, OECD Series on Testing and Assessment. OECD Publishing, Paris. <https://doi.org/10.1787/9789264274822-en>
- OECD, 2014. No. 203: New Guidance Document on an Integrated Approach on Testing and Assessment (IATA) for Skin Corrosion and Irritation, OECD Series on Testing and Assessment. OECD Publishing, Paris.
- OECD, 2007. No. 69: Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)SAR] Models, OECD Series on Testing and Assessment. OECD Publishing, Paris. <https://doi.org/10.1787/9789264085442-en>.
- OECD, 2005. No. 34: Guidance Document on the Validation and International Acceptance of New or Updated Test Methods for Hazard Assessment, OECD Environment, Health and Safety Publications Series on Testing and Assessment. OECD Publishing, Paris.
- OECD, 1998. No. 1: OECD Principles of Good Laboratory Practice, Series on Principles of Good Laboratory Practice and Compliance Monitoring. OECD Publishing, Paris.
- Pamies, D., Leist, M., Coecke, S., Bowe, G., Allen, D.G., Gstraunthaler, G., Bal-Price, A., Pistollato, F., Vries, R.B.M. de, Hogberg, H.T., Hartung, T., Stacey, G., 2022. Guidance document on Good Cell and Tissue Culture Practice 2.0 (GCCP 2.0). ALTEX - Alternatives to animal experimentation 39, 30–70.
<https://doi.org/10.14573/altex.2111011>
- Parish, S.T., Aschner, M., Casey, W., Corvaro, M., Embry, M.R., Fitzpatrick, S., Kidd, D., Kleinstreuer, N.C., Lima, B.S., Settivari, R.S., Wolf, D.C., Yamazaki, D., Boobis, A., 2020. An evaluation framework for new approach methodologies (NAMs) for human health safety assessment. Regulatory Toxicology and Pharmacology 112, 104592.
<https://doi.org/10.1016/j.yrtph.2020.104592>

- Petersen, E., 2021. Characteristics to consider when selecting a positive control material for an in vitro assay. *ALTEX*. <https://doi.org/10.14573/altex.2102111>
- Petersen, E.J., Ceger, P., Allen, D.G., Coyle, J., Derk, R., Garcia-Reyero, N., Gordon, J., Kleinstreuer, N.C., Matheson, J., McShan, D., Nelson, B.C., Patri, A.K., Rice, P., Rojanasakul, L., Sasidharan, A., Scarano, L., Chang, X., 2022a. U.S. federal agency interests and key considerations for new approach methodologies for nanomaterials. *ALTEX - Alternatives to animal experimentation* 39, 183–206. <https://doi.org/10.14573/altex.2105041>
- Petersen, E.J., Elliott, J.T., Gordon, J., Kleinstreuer, N.C., Reinke, E., Roesslein, M., Toman, B., 2022b. Technical framework for enabling high-quality measurements in new approach methodologies (NAMs). *ALTEX - Alternatives to animal experimentation*. <https://doi.org/10.14573/altex.2205081>
- Petersen, E.J., Uhl, R., Toman, B., Elliott, J.T., Strickland, J., Truax, J., Gordon, J., 2022c. Development of a 96-well electrophilic allergen screening assay for skin sensitization using a measurement science approach. *Toxics* 10, 257. <https://doi.org/10.3390/toxics10050257>
- Pham, L.L., Watford, S., Pradeep, P., Martin, M.T., Thomas, R., Judson, R., Setzer, R.W., Paul Friedman, K., 2020. Variability in in vivo studies: Defining the upper limit of performance for predictions of systemic effect levels. *Comput Toxicol* 15, 1–100126. <https://doi.org/10.1016/j.comtox.2020.100126>
- Piersma, A.H., van Benthem, J., Ezendam, J., Kienhuis, A.S., 2018. Validation redefined. *Toxicol In Vitro* 46, 163–165. <https://doi.org/10.1016/j.tiv.2017.10.013>
- Prior, H., Casey, W., Kimber, I., Whelan, M., Sewell, F., 2019. Reflections on the progress towards non-animal methods for acute toxicity testing of chemicals. *Regul Toxicol Pharmacol* 102, 30–33. <https://doi.org/10.1016/j.yrtph.2018.12.008>
- Rooney, J.P., Choksi, N.Y., Ceger, P., Daniel, A.B., Truax, J., Allen, D., Kleinstreuer, N., 2021. Analysis of variability in the rabbit skin irritation assay. *Regulatory Toxicology and Pharmacology* 122, 104920. <https://doi.org/10.1016/j.yrtph.2021.104920>
- Rösslein, M., Elliott, J.T., Salit, M., Petersen, E.J., Hirsch, C., Krug, H.F., Wick, P., 2015. Use of Cause-and-Effect Analysis to Design a High-Quality Nanocytotoxicology Assay. *Chem. Res. Toxicol.* 28, 21–30. <https://doi.org/10.1021/tx500327y>
- Rusyn, I., Chiu, W.A., Wright, F.A., 2022. Model systems and organisms for addressing inter- and intra-species variability in risk assessment. *Regul Toxicol Pharmacol* 132, 105197. <https://doi.org/10.1016/j.yrtph.2022.105197>

- Sewell, F., Doe, J., Gellatly, N., Ragan, I., Burden, N., 2017. Steps towards the international regulatory acceptance of non-animal methodology in safety assessment. *Regul Toxicol Pharmacol* 89, 50–56. <https://doi.org/10.1016/j.yrtph.2017.07.001>
- Shaffer, R.M., 2021. Environmental Health Risk Assessment in the Federal Government: A Visual Overview and a Renewed Call for Coordination. *Environ. Sci. Technol.* 55, 10923–10927. <https://doi.org/10.1021/acs.est.1c01955>
- Slezák, P., Waczulíková, I., 2011. Letter to the Editor: Reproducibility and Repeatability. *Physiol Res* 60, 203–205.
- Smirnova, L., Hogberg, H.T., Leist, M., Hartung, T., 2014. Developmental neurotoxicity - challenges in the 21st century and in vitro opportunities. *ALTEX* 31, 129–156. <https://doi.org/10.14573/altex.1403271>
- Strickland, J., Clippinger, A.J., Brown, J., Allen, D., Jacobs, A., Matheson, J., Lowit, A., Reinke, E.N., Johnson, M.S., Quinn, M.J., Mattie, D., Fitzpatrick, S.C., Ahir, S., Kleinstreuer, N., Casey, W., 2018. Status of acute systemic toxicity testing requirements and data uses by U.S. regulatory agencies. *Regul Toxicol Pharmacol* 94, 183–196. <https://doi.org/10.1016/j.yrtph.2018.01.022>
- Strickland, J., Daniel, A.B., Allen, D., Aguila, C., Ahir, S., Bancos, S., Craig, E., Germolec, D., Ghosh, C., Hudson, N.L., Jacobs, A., Lehmann, D.M., Matheson, J., Reinke, E.N., Sadrieh, N., Vukmanovic, S., Kleinstreuer, N., 2019. Skin sensitization testing needs and data uses by US regulatory and research agencies. *Arch Toxicol* 93, 273–291. <https://doi.org/10.1007/s00204-018-2341-6>
- Thomas, R.S., Bahadori, T., Buckley, T.J., Cowden, J., Deisenroth, C., Dionisio, K.L., Frithsen, J.B., Grulke, C.M., Gwinn, M.R., Harrill, J.A., Higuchi, M., Houck, K.A., Hughes, M.F., Hunter, E.S., Isaacs, K.K., Judson, R.S., Knudsen, T.B., Lambert, J.C., Linnenbrink, M., Martin, T.M., Newton, S.R., Padilla, S., Patlewicz, G., Paul-Friedman, K., Phillips, K.A., Richard, A.M., Sams, R., Shafer, T.J., Setzer, R.W., Shah, I., Simmons, J.E., Simmons, S.O., Singh, A., Sobus, J.R., Strynar, M., Swank, A., Tornero-Valez, R., Ulrich, E.M., Villeneuve, D.L., Wambaugh, J.F., Wetmore, B.A., Williams, A.J., 2019. The next generation blueprint of computational toxicology at the U.S. Environmental Protection Agency. *Toxicol Sci* 169, 317–332. <https://doi.org/10.1093/toxsci/kfz058>
- Tox21, n.d. Tox21 - Toxicology in the 21st Century [WWW Document]. Tox21. URL <https://tox21.gov/> (accessed 5.9.23).
- Tsuji, R., Crofton, K.M., 2012. Developmental neurotoxicity guideline study: issues with methodology, evaluation and regulation. *Congenit Anom (Kyoto)* 52, 122–128. <https://doi.org/10.1111/j.1741-4520.2012.00374.x>
- van der Zalm, A.J., Barroso, J., Browne, P., Casey, W., Gordon, J., Henry, T.R., Kleinstreuer, N.C., Lowit, A.B., Perron, M., Clippinger, A.J., 2022. A framework for establishing

- scientific confidence in new approach methodologies. *Arch Toxicol* 96, 2865–2879.
<https://doi.org/10.1007/s00204-022-03365-4>
- Wikoff, D., Lewis, R.J., Erraguntla, N., Franzen, A., Foreman, J., 2020. Facilitation of risk assessment with evidence-based methods – A framework for use of systematic mapping and systematic reviews in determining hazard, developing toxicity values, and characterizing uncertainty. *Regulatory Toxicology and Pharmacology* 118, 104790.
<https://doi.org/10.1016/j.yrtph.2020.104790>
- Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018. <https://doi.org/10.1038/sdata.2016.18>
- Wolffe, T.A.M., Whaley, P., Halsall, C., Rooney, A.A., Walker, V.R., 2019. Systematic evidence maps as a novel tool to support evidence-based decision-making in chemicals policy and risk management. *Environ Int* 130, 104871.
<https://doi.org/10.1016/j.envint.2019.05.065>
- Zhang, J.H., Chung, T.D., Oldenburg, K.R., 1999. A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J Biomol Screen* 4, 67–73.
<https://doi.org/10.1177/108705719900400206>
- Zhang, X.D., 2011. Illustration of SSMD, z score, SSMD*, z* score, and t statistic for hit selection in RNAi high-throughput screens. *J Biomol Screen* 16, 775–785.
<https://doi.org/10.1177/1087057111405851>

APÉNDICE A: GLOSARIO

DESCARGO DE RESPONSABILIDAD: Este glosario está destinado a respaldar el uso de terminología en este informe. El significado o uso de un término en particular puede variar según las agencias reguladoras.

3Rs: Principios de la técnica experimental humana, específicamente con respecto a la reducción (es decir, minimizar el número de animales utilizados), reemplazo (es decir, uso de una especie de orden inferior o prueba no animal), o refinamiento (es decir, minimizar el dolor o angustia que los animales de investigación podrían soportar debido a una técnica particular) del uso de animales en investigación y pruebas de seguridad química.

Adecuado para el propósito: Hacer coincidir el tipo y certeza de la información proporcionada por una MNE (o conjunto de MNE) con el tipo y certeza de información necesaria para una decisión determinada (EPA, 2021a).

Base de datos/lista curada: Es un conjunto estructurado de información bien caracterizada y confiable, que se compila y conserva cuidadosamente.

Buenas Prácticas de Cultivo Celular (BPCC): Conjunto de principios desarrollados para uso práctico en el laboratorio, para asegurar la reproducibilidad del trabajo *in vitro* (basado en células) y mejorar la calidad de los datos científicos (Pamies et al., 2022).

Buenas Prácticas de Laboratorio (BPL): Reglamentos promulgados por autoridades como la EPA de los Estados Unidos, la FDA de los Estados Unidos y la OCDE que describen los procedimientos de mantenimiento de registros y garantía de calidad para los registros de laboratorio que serán la base para la presentación de datos a las agencias reguladoras nacionales.

Buenas Prácticas de Métodos In Vitro (BPMIV): Marco integral descrito en el Documento de Orientación de la OCDE sobre Buenas Prácticas de Métodos In Vitro (OCDE, 2018), que proporciona recomendaciones para el desarrollo, validación, aceptación o calificación regulatoria y uso, de métodos *in vitro*.

Calificación: Conclusión de que se puede confiar en los resultados de una evaluación utilizando un modelo o ensayo validado, para tener una interpretación y aplicación específicas en el desarrollo de productos y la toma de decisiones regulatorias.

Clasificación de peligros: Asignación de un peligro por una sustancia o producto químico a una categoría, con base en los resultados de un método de prueba estándar de un criterio de valoración de toxicidad específico; más comúnmente utilizado para fines de etiquetado.

Compuestos de referencia: Sustancias químicas seleccionadas para su uso durante la investigación, desarrollo o evaluación de un método de prueba propuesto, debido a que se conoce

su respuesta en el método de prueba de referencia o en la especie de interés (véase “método de prueba de referencia”).

Concordancia: Es la cercanía de acuerdo o consistencia entre dos variables. La concordancia puede usarse para describir cualitativamente la relevancia biológica de una MNE en comparación con el método de prueba en animales de referencia y/o los datos de referencia en seres humanos, cuando los hay disponibles. La concordancia también se puede utilizar para describir cuantitativamente la proporción de todas las sustancias químicas analizadas que se clasifican correctamente como positivas o negativas y el término a menudo se usa indistintamente con “exactitud” en este contexto.

Confiabilidad: Medida del grado en el que un método de prueba se puede realizar de manera reproducible dentro y entre laboratorios, a lo largo del tiempo. Se evalúa calculando la reproducibilidad intra- e interlaboratorios y la repetibilidad intralaboratorio.

Contexto de uso (CDU): Es una descripción claramente articulada que delinea la manera y propósito de uso de un método, enfoque o aplicación particular.

Control de calidad: Conjunto de actividades o muestras que se utilizan para verificar que la calidad del producto o método se mantenga según lo previsto.

Criterio de valoración: Es el proceso biológico o químico, respuesta, o efecto, evaluado por un método de prueba.

Curva estándar: Método cuantitativo de graficación de datos de ensayo para determinar la concentración de una sustancia en una muestra desconocida, comparando dicha muestra con una muestra estándar de concentración conocida (a menudo con el uso del material de control positivo).

Desarrollador de métodos de prueba: Organización o individuo que inicialmente diseña un método de prueba y asegura su reproducibilidad e idoneidad para el uso previsto.

Desempeño: Son las características de exactitud y confiabilidad de un método de prueba (véanse “exactitud”, “confiabilidad”).

Discordancia: Es la proporción de todas las sustancias químicas analizadas que fueron incorrectamente clasificados como positivos o negativos.

Dominio de aplicabilidad: Los tipos de sustancias químicas que se pueden probar utilizando un método, o los tipos de sustancias químicas para las cuales se consideran aceptables los resultados producidos por ese método.

Enfoque definido (ED): Consiste en datos de entrada generados con un conjunto definido de fuentes de información y un procedimiento fijo de interpretación de datos, para obtener un resultado que puede ser usado por sí mismo o junto con otras fuentes de información dentro de una IATA, para satisfacer una necesidad regulatoria específica. Se puede utilizar un enfoque

definido para las pruebas y la evaluación para respaldar la identificación de peligros, la caracterización de peligros y/o la evaluación de la seguridad de sustancias químicas (OCDE, 2017).

Enfoque Integrado para Pruebas y Evaluaciones (IATA): Enfoque basado en múltiples fuentes de información utilizadas para la identificación de peligros, caracterización de peligros y/o evaluación de seguridad de sustancias químicas (OCDE, 2017).

Ensayo: El sistema experimental utilizado. A menudo se usa indistintamente con “prueba” y “método de prueba”.

Especie objetivo: Especie para la cual se busca información sobre la toxicidad potencial de una sustancia química.

Especificidad: Proporción de todas las sustancias químicas negativas clasificadas correctamente como negativas en un método de prueba. Es una medida de la exactitud del método de prueba. Este término también se utiliza para describir la capacidad de un método analítico para detectar un analito específico.

Estabilidad: Capacidad de un material de ensayo (por ejemplo, una sustancia de prueba, aparato de ensayo, reactivo o analito), para producir resultados similares y aceptables durante un período de tiempo en un ambiente determinado.

Estándares de desempeño: Estándares, basados en un método de prueba validado, que proporcionan una base para evaluar la comparabilidad de un método de prueba propuesto mecánicamente y funcionalmente similar. Se incluyen (1) componentes esenciales del método de prueba; (2) una lista de sustancias químicas de referencia seleccionadas de entre las sustancias químicas utilizadas para demostrar el desempeño aceptable del método de prueba validado; y (3) los niveles comparables de exactitud y confiabilidad, con base en lo que se obtuvo para el método de prueba validado, que el método de prueba propuesto debe demostrar cuando se evalúa utilizando la lista mínima de sustancias químicas de referencia.

Evaluación de riesgos: Evaluación de los posibles efectos adversos para la salud y el medio ambiente, de una especie objetivo a partir de exposiciones a agentes exógenos.

Exactitud equilibrada: Es una métrica estadística empleada para dar cuenta de un conjunto de datos desequilibrado, en donde una “clase” aparece mucho más que la otra (por ejemplo, mayor número de negativos que positivos). La exactitud equilibrada se calcula como la media aritmética de la sensibilidad y especificidad.

Exactitud: La cercanía de concordancia entre un resultado del método de prueba y un valor de referencia aceptado.

Falso negativo: Sustancia identificada incorrectamente como negativa por un método de prueba, en relación con los datos de referencia especificados.

Falso positivo: Es una sustancia identificada incorrectamente como positiva por un método de prueba, en relación con los datos de referencia especificados.

Identificación del peligro: Parte de la evaluación de riesgos asociada con la determinación de si la exposición a una sustancia en particular está o podría estar asociada con efectos adversos para la salud o adversos ecológicos.

Método de prueba de referencia: Método de prueba aceptado que se emplea con fines regulatorios para evaluar el potencial de una sustancia de prueba de ser peligrosa para la especie de interés.

Método de prueba validado: Método de prueba aceptado para el cual se han realizado estudios de validación, para determinar la exactitud, confiabilidad y relevancia de dicho método para un uso específico propuesto.

Método de prueba: Proceso o procedimiento utilizado para obtener información sobre las características de una sustancia o agente. Los métodos de pruebas toxicológicas generan información referente a la capacidad de una sustancia o agente para producir un efecto biológico especificado, bajo condiciones especificadas. Se puede usar indistintamente con “test” y “assay” (prueba y ensayo). Véase también “método de prueba validado” y “método de prueba de referencia”.

Metodología de nuevo enfoque (MNE): Constituye una referencia ampliamente descriptiva a cualquier tecnología, metodología, enfoque o combinación de los mismos, que pueda utilizarse para proporcionar información sobre el peligro químico y la evaluación de riesgos, y que respalda el reemplazo, reducción o refinamiento (3Rs) del uso de animales.

Órgano objetivo: Órgano para el que se busca información sobre la toxicidad potencial de una sustancia química.

Patrocinador del método de prueba: Organización o individuo que presenta un método de prueba para su consideración; esta misma organización o individuo también puede ser el desarrollador del método de prueba.

Peligro: Es el potencial de un efecto adverso o nocivo para la salud o nocivo ecológico.

Potencia: Medida de la actividad biológica o química relativa de una sustancia. La potencia de una sustancia sola puede diferir a causa de diferentes efectos biológicos o bioquímicos.

Precisión: Es la cercanía de las mediciones individuales de un analito después de múltiples análisis, a menudo de una sola muestra. La precisión a menudo se expresa como el coeficiente de variación.

Predictividad negativa: Es la proporción de respuestas negativas correctas en relación con datos de referencia definidos, entre las sustancias que dan negativo por un método de prueba. Es un indicador de la exactitud del método de prueba. La predictividad negativa es una función de la

especificidad del método de prueba y de la prevalencia de negativos entre las sustancias probadas.

Predictividad positiva: Es la proporción de respuestas positivas correctas entre las sustancias que dan positivo, en relación con los datos de referencia definidos por un método de prueba. Es un indicador de la exactitud del método de prueba. La predictividad positiva es una función de la sensibilidad del método de prueba y la prevalencia de positivos entre las sustancias probadas.

Presentación del método de prueba: Compendio de documentación justificativa de un método de prueba propuesto para aplicación regulatoria u otra aplicación definida. Una presentación de método de prueba generalmente incluye registros de estudios de validación que se han completado para caracterizar la utilidad y limitaciones del método de prueba, para un requisito o aplicación de prueba reglamentaria específica propuesta, así como otra documentación adecuada sobre la validez científica; todo esto se prepara de conformidad con los lineamientos de presentación de métodos de prueba.

Procedimientos normalizados de operación (PNO): Procedimientos formales por escrito que describen cómo se deben realizar operaciones específicas del laboratorio. Estos son requeridos por los lineamientos de las BPL.

Protocolo: Es la descripción precisa paso a paso de un método de prueba, incluyendo la lista de todos los reactivos necesarios y todos los criterios y procedimientos para generar y evaluar los datos de la prueba.

Prueba: Sistema experimental utilizado; puede emplearse indistintamente con “método de prueba” y “ensayo”.

Recuperación: Método cuantitativo para verificar la eficiencia y reproducibilidad de un método de extracción, mediante una comparación de los resultados de las muestras extraídas, con aquellos de muestras sembradas de una matriz similar y/o blancos sembrados.

Relevancia biológica: Es una medida de idoneidad para evaluar los efectos de una sustancia química dentro de los taxones de interés.

Relevancia mecanicista: Medida de idoneidad para evaluar el proceso bioquímico o ruta por la cual una sustancia química puede ejercer un efecto.

Repetibilidad: Es la consistencia de los resultados de la prueba obtenidos cuando el procedimiento se realiza sobre la misma sustancia en condiciones idénticas, en un periodo de tiempo determinado; “... la cercanía de coincidencia entre resultados independientes obtenidos con el mismo método en uno o más sujetos idénticos (u objetos o materiales de prueba), *bajo las mismas condiciones*” (Slezák y Waczulíková, 2011).

Reproducibilidad interlaboratorios: Es una medida de si diferentes laboratorios calificados que utilizan el mismo protocolo y productos químicos de prueba, pueden producir resultados cualitativa y cuantitativamente similares. La reproducibilidad interlaboratorios se determina

durante los procesos de prevalidación y validación e indica la medida en que un método de prueba puede ser transferido con éxito entre laboratorios.

Reproducibilidad: Es la consistencia de los resultados de pruebas individuales obtenidos utilizando el mismo protocolo y muestras de prueba; “... la cercanía del acuerdo entre resultados independientes obtenidos con el mismo método sobre uno o más sujetos idénticos (u objetos, o materiales de prueba), pero *bajo condiciones diferentes* (diferentes observadores, laboratorios etc.)” (Slezák y Waczulíková, 2011).

Robustez: Capacidad de un método de ser reproducido bajo diferentes condiciones o circunstancias, sin presentar diferencias inesperadas en los resultados obtenidos.

Ruta de resultado adverso (RRA): Es una representación estructurada de eventos secuenciales que ocurren a diferentes niveles de organización biológica, dando como resultado un efecto adverso cuando un organismo es expuesto a una sustancia.

Sensibilidad: Proporción de todas las sustancias químicas positivas clasificadas correctamente como positivas en un método de prueba. La sensibilidad también puede definirse en el contexto de los límites de detección, como la concentración más baja de analito que se puede medir con una exactitud y/o precisión aceptables.

Transferibilidad: Capacidad de un método o procedimiento de prueba para ser ejecutado de manera precisa y confiable en diferentes laboratorios competentes. Véase también “reproducibilidad interlaboratorios”.

Validación: Proceso por el cual se establece la exactitud, confiabilidad y relevancia de un procedimiento para un propósito específico. La validación para un propósito específico no implica la validación para otros fines específicos. Puede ser necesaria una calificación adicional para un contexto de uso particular.

APÉNDICE B: HERRAMIENTAS DE CALIDAD

1.0 Diagramas de flujo

En un diagrama de flujo, cada paso de un protocolo se incluye en el diagrama (consulte la Figura S1 para ver un ejemplo). Esto puede ser útil en el diseño de medidas de control para el experimento al garantizar que cada paso se cubra y pueda ser rastreado cuando sea factible. Es posible que más de una medición de control abarque un solo paso y, a la inversa, una medición de control puede abarcar múltiples pasos. Además, la comparación de las medidas de control entre MNE puede revelar pasos similares entre diferentes MNE. Para estos pasos, es probable que las fuentes de variabilidad también serán similares.

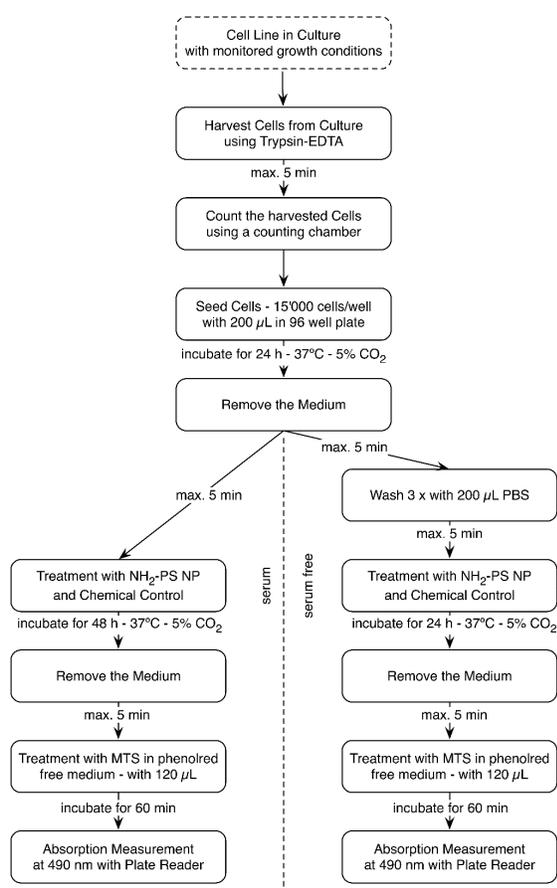


Figura suplementaria 1. Diagrama de flujo que describe el protocolo del MTS (3-(4,5-dimetiltiazol-2-il)-5-(3-carboximetoxifenil)-2-(4-sulfofenil)-2H-tetrazolio), modificado para un estudio interlaboratorios. Esta figura fue reimpressa con permiso de Elliott et al. (2017).

2.0 Análisis de Causa-Efecto

El análisis de causa y efecto (C&E) es una herramienta conceptual que se puede emplear para identificar posibles fuentes clave de variabilidad y mostrarlas utilizando diagramas de C&E, en ocasiones conocidos como diagramas de espina de pescado (consulte la Figura S2 para ver un ejemplo). El proceso de desarrollo de estos diagramas puede incluir una lluvia de ideas y la revisión de literatura sobre un ensayo. Cada rama del diagrama de C&E indica una fuente clave de variabilidad esperada. Los diagramas de C&E también pueden respaldar el desarrollo de nuevas MNE, porque las ramas compartidas en diagramas de C&E (por ejemplo, realizar mediciones usando el mismo tipo de ensayo de citotoxicidad) se pueden hacer más rápidamente, es probable que requieran medidas de control similares y posiblemente tendrán estrategias similares de mitigación de variabilidad. El análisis de los diagramas de C&E puede ayudar a identificar aspectos de un método que pueden ser difíciles de estandarizar (por ejemplo, un instrumento difícil de calibrar o reactivos de ensayo inestables). Los diagramas de C&E pueden ayudar a guiar las pruebas de robustez y la selección de medidas de control para que, de manera ideal, se analicen las fuentes de variabilidad en cada rama y subrama del diagrama de C&E.

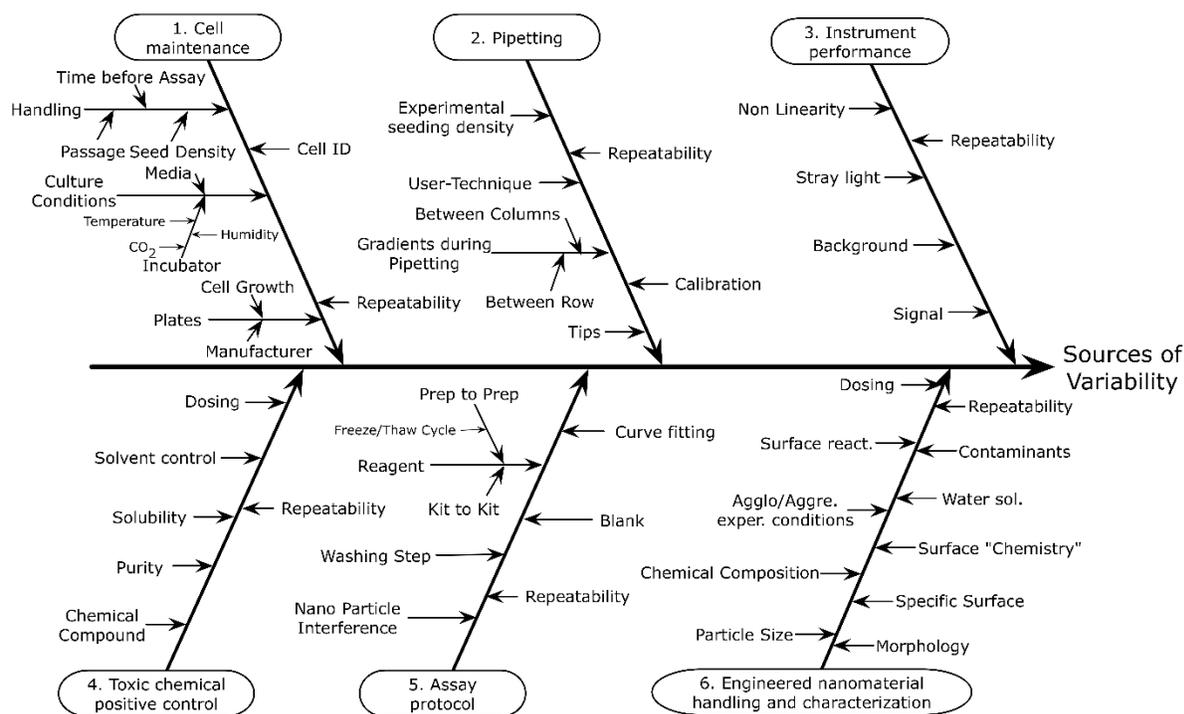


Figura Suplementaria 2. Diagrama de causa-efecto de un ensayo de citotoxicidad del MTS diseñado para usarse con nanomateriales diseñados por ingeniería. Reimpreso con permiso de Rösslein et al. (2015).

3.0 Gráficas de control

Las medidas de control pueden ser experimentos preliminares de ejecución única (por ejemplo, para probar posibles sesgos), mediciones periódicas realizadas a una frecuencia predeterminada (por ejemplo, evaluación del desempeño del instrumento y calibración del mismo), y medidas de control en proceso realizadas cada vez que se ejecuta el ensayo. En algunos ensayos que permiten analizar un número limitado de muestras al mismo tiempo (por ejemplo, un ensayo de inhalación con exposición a productos químicos en aerosol utilizando un sistema de exposición de flujo directo), las mediciones periódicas pueden ser útiles. Los experimentos preliminares de ejecución única pueden ser útiles para evaluar si una sustancia de prueba puede causar un sesgo, por ejemplo, si una nanopartícula puede adsorber un reactivo de ensayo clave. Las medidas de control en proceso se pueden emplear para medir fuentes clave de variabilidad cada vez que se realiza el ensayo. Por ejemplo, una medición de control en proceso común es el control positivo. Esta medición de control puede revelar si se alcanza una respuesta máxima (efecto del 100%) y también puede usarse para demostrar la sensibilidad de la respuesta del ensayo a concentraciones químicas que producen respuestas más bajas y moderadas. Las consideraciones clave para seleccionar un control positivo han sido descritas por Petersen et al. (2021). Algunas mediciones adicionales comunes de control en proceso, son un control sin células y sin reactivos de ensayo adicionales y un control con células y reactivos de ensayo adicionales añadidos, pero sin sustancias de prueba. Podría no ser posible incluir todas las potenciales medidas de control en el proceso y puede haber compensaciones en términos de qué medidas de control en proceso incluir.

Las fuentes clave de variabilidad pueden irse monitoreando a lo largo del tiempo y entre experimentos, mediante el uso de gráficas de control que muestran los valores medios y la variabilidad de las medidas de control en proceso (consulte la Figura S3 para ver un ejemplo). Esto puede ser útil para evaluar si hay cambios sistemáticos en los valores medios o de variabilidad a lo largo del tiempo, que sugieran que algo en el ensayo podría haber cambiado (por ejemplo, inestabilidad de un reactivo). Revisar las hojas de verificación puede ser útil para identificar por qué se produjeron cambios.

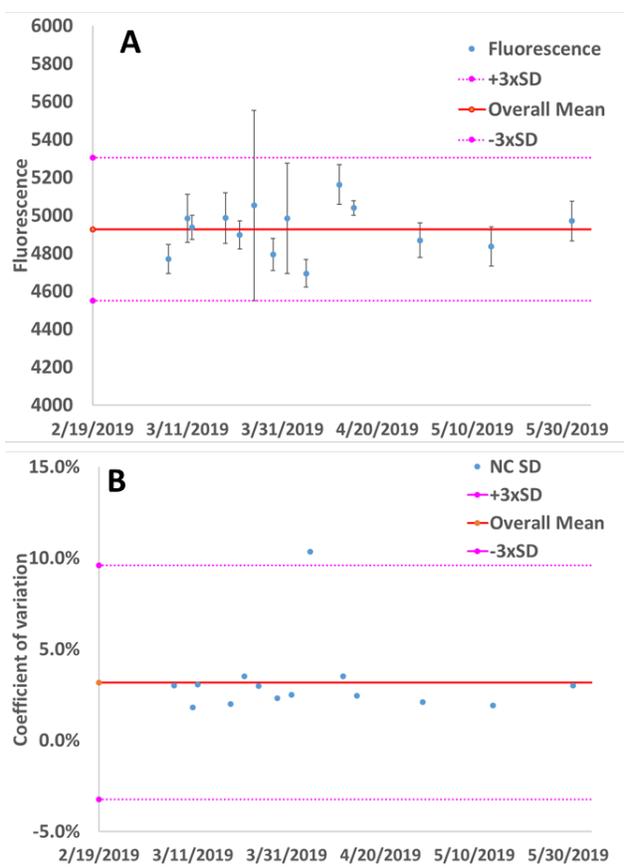


Figura suplementaria 3. Datos de la gráfica de control del ensayo de detección de alérgenos electrofílicos, método de fluorescencia, para el control negativo de incubadora: (A) media y (B) coeficiente de variación de todos los experimentos, dependiendo de la fecha en que se realizaron. En la gráfica B, un valor es atípico para el coeficiente de variación y está fuera de las especificaciones de este estudio (media general ± 3 veces el valor promedio de la desviación estándar). Además, no existe una tendencia sistemática con los valores de la media o el coeficiente de variación a lo largo del tiempo. Esta figura fue modificada y reimpressa con permiso de Petersen et al. (2022c), mientras que las leyendas de figura fueron modificadas y reimpresas de Petersen et al. (2022b).

4.0 Hojas de Verificación

Para registrar metadatos clave y cumplir con los requisitos de las BPL, se pueden usar hojas de verificación. Si no se aplican las BPL cuando se está desarrollando un método, los principios establecidos en ellas, como el registro de datos clave en hojas de verificación, serán importantes si las MNE están destinados a presentación regulatoria, en donde es posible que se requiera el cumplimiento de las BPL. Algunos recursos alternativos relevantes durante el desarrollo del ensayo son el documento de orientación sobre BPCC 2.0 (Pamies et al., 2022) y GIVIMP (OCDE, 2018), que son específicos para el trabajo basado en células, con listas adicionales de gestión de calidad para garantizar la reproducibilidad y datos científicos de alta calidad. Las hojas de verificación también pueden ser útiles para monitorear las medidas de control en proceso y apoyar la solución de problemas cuando éstos surjan. Por ejemplo, se pueden usar para documentar cambios en los números de lote de reactivos y suministros consumibles (por ejemplo, pipetas y microplacas). Esto puede ser útil para solucionar problemas si una especificación no se cumple con más frecuencia que otras. También es posible que las hojas de verificación que incluyen calculadoras de datos evalúen si los datos de una ejecución de ensayo cumplen con todas las especificaciones y efectúen evaluaciones estadísticas de los resultados de dicho ensayo. Se pueden usar cuadernos electrónicos y sistemas de gestión de información de laboratorio en lugar de hojas de verificación escritas (con ventajas como la facilidad de almacenamiento, transferibilidad entre laboratorios y posibilidades de búsqueda).

5.0 Gráficas de Dispersión

Las gráficas de dispersión se pueden usar para evaluar si existe una interacción entre diferentes medidas de control en proceso o entre esas medidas de control y los resultados de la sustancia de prueba (consulte la Figura S4 para ver un ejemplo). Los resultados de la sustancia de prueba deben ser independientes de los resultados de las medidas de control en proceso, dentro del rango establecido por las especificaciones.

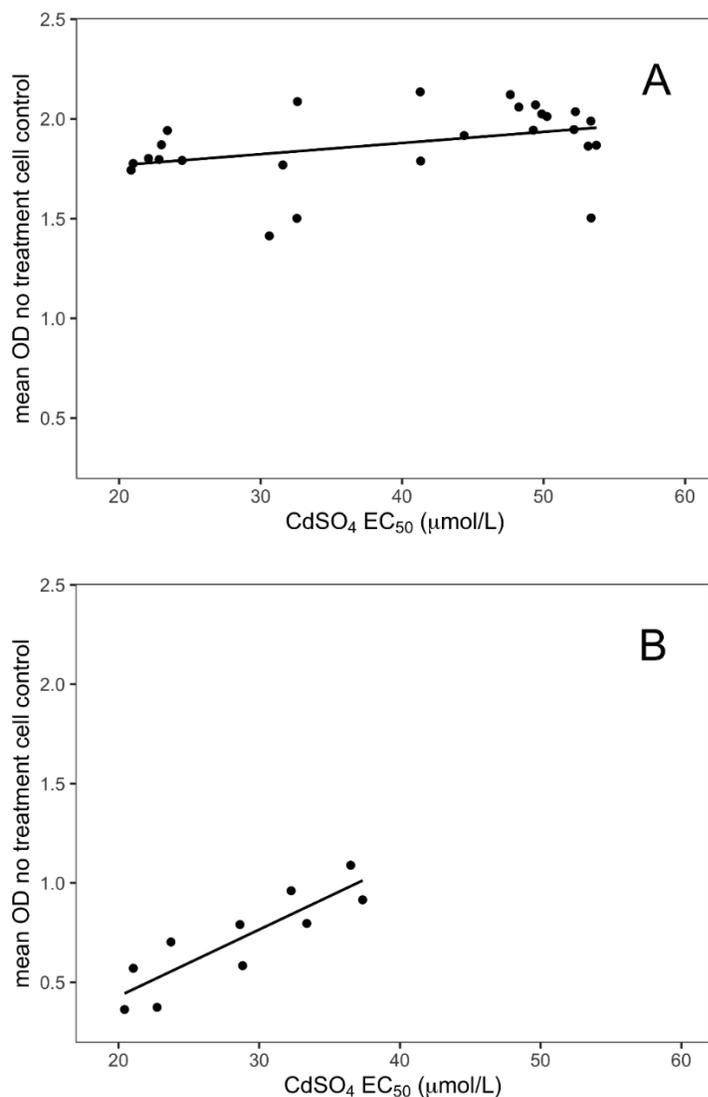


Figura complementaria 4. Correlación de los valores de CE_{50} del sulfato de cadmio ($CdSO_4$) determinados usando el ensayo MTS con valores medios de densidad óptica (DO) de células A549 no expuestas a la sustancia de prueba o al control positivo. Estos datos muestran una falta de interacción entre los valores de CE_{50} (parte A) o una interacción (parte B) dependiendo del rango de valores medios de DO. Las líneas continuas son ajustes de regresión lineal. La pendiente en la parte B es estadísticamente diferente de 0, lo que indica que el valor CE_{50} está correlacionado con los valores de DO. Modificado y reimpresso con permiso de Petersen et al. (2022b) y las leyendas de figura fueron reimpresas con permiso de Petersen et al. (2022b).

APÉNDICE C: EVALUACIÓN DE MÉTODOS ANALÍTICOS

La información del Apéndice C proviene de varias referencias, entre ellas “Validación y Aceptación Regulatoria de Métodos de Pruebas Toxicológicas” (ICCVAM, 1997), “Guía de Validación de Métodos Bioanalíticos para la Industria” (FDA, 2018), GD 34 de la OCDE (OCDE, 2005), y el documento GIVIMP de la OCDE (OCDE, 2018). La información contenida en el Apéndice C tiene por objeto proporcionar un marco general sobre la evaluación de métodos. Algunos aspectos pueden o no ser aplicables a todos los métodos.

1.0 Límites de Detección y Cuantificación

Los límites de detección y cuantificación definen el rango de un método analítico. El método analítico a menudo se prueba ya sea en el límite inferior de cuantificación o en el límite de detección, cualquiera que sea más apropiado para el sistema. Este proceso determina el extremo inferior del rango de un método analítico estableciendo la menor cantidad o concentración de analito que puede ser detectado de manera confiable (límite de detección) o cuantificado (límite inferior de cuantificación) por encima del blanco de reactivos o dentro de la curva estándar. La concentración más alta en la curva estándar determina el límite superior de cuantificación.

2.0 Identificación de Interferencias

Los desarrolladores de métodos deben documentar las sustancias interferentes que pueden provenir de componentes críticos y no críticos del método, incluyendo cualquier interferencia con la señal detectada (por ejemplo, fluorescencia/absorbancia, luciferasa, enzimática) del método. La interferencia también puede provenir de consumibles, tales como ciertos plásticos en métodos de prueba de perturbadores endocrinos, así como otros reactivos en el método. Las sustancias potencialmente interferentes incluyen, pero no se limitan a, componentes de matriz endógenos, metabolitos, productos de descomposición y otros xenobióticos. Si el método está destinado a cuantificar más de un analito, todos los analitos previstos deberán ser probados para asegurar que no haya interferencia. A menudo se emplean muestras blanco para probar la interferencia.

3.0 Evaluación de la Precisión Analítica

Es útil incluir una descripción de la precisión del método analítico utilizado y cualesquiera otras pruebas de precisión, tales como las que evalúan la variabilidad del desempeño, cuando diferentes personas emplean el método propuesto, o cuando se utiliza instrumentación diferente para el método, así como la participación en estudios de comparación interlaboratorios, cuando sea posible.

4.0 Estabilidad de los Materiales

La estabilidad en lo relacionado con los materiales utilizados en la MNE (por ejemplo, sustancias de prueba, aparatos de ensayo, reactivos y analitos), se refiere a la capacidad de un material para producir resultados similares y aceptables durante un período de tiempo en un ambiente

determinado. La estabilidad de la MNE misma se discutió anteriormente (Apéndice B, Sección 3.0) con la evaluación de las gráficas de control a lo largo del tiempo. Se deben evaluar los efectos de las condiciones de toma de muestras, del manejo y almacenamiento de las mismas. A los materiales que pueden verse afectados por problemas de estabilidad se les deberán asignar fechas de caducidad apropiadas. Debe garantizarse la estabilidad de la sustancia de prueba, reactivos y aparatos de ensayo (por ejemplo, las microplacas de plástico) para evitar interferencias de los productos de degradación y cambios en la dosis real aplicada. Los estudios de estabilidad que se realicen sobre cualquiera de los componentes deben incluirse en los informes. Se debe evaluar la estabilidad de cualquier mezcla química o muestra preparada antes del día del estudio antes de su uso en estudios de desarrollo, validación o calificación.

La estabilidad química de una mezcla o matriz dada bajo condiciones específicas, para intervalos de tiempo determinados, se evalúa de varias maneras. Las evaluaciones de estabilidad previas al estudio deben cubrir las condiciones esperadas de manejo y almacenamiento de la muestra durante la realización del estudio, incluyendo las condiciones en el sitio de prueba, durante el envío y en todos los demás sitios secundarios. La estabilidad de un analito en un sistema particular de mezcla, matriz y contenedor es relevante solo para ese sistema de mezcla, matriz y contenedor, y no deberá extrapolarse a otros sistemas. Las pruebas de estabilidad deben evaluar la estabilidad de los analitos para el almacenamiento a largo plazo (congelados a la temperatura de almacenamiento prevista) y a corto plazo (en mesa, temperatura ambiente) y luego de ciclos de congelación y descongelación y del proceso analítico. Las condiciones utilizadas en los experimentos de estabilidad deben reflejar situaciones que puedan encontrarse durante el manejo real y análisis de la muestra. Si durante el análisis de la muestra en un estudio las condiciones de almacenamiento cambiaron y/o excedieron las condiciones de almacenamiento de la muestra evaluadas durante la caracterización técnica del método, se deberá establecer la estabilidad bajo estas nuevas condiciones. Se deben realizar estudios de estabilidad independientes y cubrir cualquier condición que pueda enfrentar un reactivo crítico. Las pruebas de estabilidad de un reactivo o material deben considerarse para una variedad de contextos, incluso durante los ciclos de congelación-descongelación y su uso en mesa de laboratorio, en soluciones stock, en muestras procesadas y en estudios de largo plazo.

5.0 Pruebas de Robustez

Robustez es la capacidad de un método de ser reproducido bajo diferentes condiciones o circunstancias, sin presentar diferencias inesperadas en los resultados obtenidos. Un aspecto de la robustez de una MNE es la consistencia de los resultados de las gráficas de control respecto al tiempo, tal como se describe en el Apéndice B, Sección 3.0. Las pruebas de robustez se utilizan a menudo para detectar cambios en los resultados provenientes de variaciones no intencionales en los reactivos o protocolos experimentales. Se recomienda la prueba de robustez para todos los aspectos de los métodos de prueba y se deben establecer rangos para todos los parámetros y mediciones siempre que sea posible.

Por ejemplo, se estableció un tiempo de incubación de 5 minutos como óptimo para un estudio, pero después de las pruebas de robustez, los datos pasaron todos los requisitos de control de

calidad a los 5 minutos más o menos 30 segundos. Por lo tanto, el tiempo de incubación aceptable de robustez probada sería de 5 minutos \pm 30 segundos.

Algunos parámetros de estudio que deben tener un rango de aceptación establecido incluyen:

- Tiempos de incubación
- Temperaturas de incubación
- pH
- Fuentes de reactivos
- Densidades de células (si es aplicable)
- Condiciones experimentales
- Software de análisis

El diseño del modelo determinará qué condiciones experimentales requerirán rangos de aceptación. Por ejemplo, los modelos con medios que fluyen necesitarán rangos de aceptación para los parámetros relacionados con el flujo. Los modelos más complejos es más probable que tengan parámetros de estudio que requieran control.

Cuando sea aplicable, también se deben informar los resultados de las pruebas de robustez de reactivos críticos y no críticos. Se deben probar diferentes proveedores (siempre que sea práctico) para determinar si se debe comprar un reactivo a un proveedor o si se pueden usar múltiples proveedores. Esto también puede ser útil para evitar interrupciones en la cadena de suministro o en caso que un fabricante deje de fabricar un producto específico.

También se deben realizar pruebas de robustez instrumental siempre que sea aplicable. La validación cruzada entre laboratorios a menudo involucra diferentes marcas de instrumentos con diferentes parámetros de desempeño y capacidades, lo cual puede sumarse a la variabilidad de los datos o cambiar parámetros del método ejecutado.

6.0 Análisis de Recuperación

Los estudios que realizan pruebas en extractos de una matriz determinada deberán llevar a cabo estudios de recuperación, para verificar la eficiencia y reproducibilidad del método de extracción. Los estudios de recuperación a menudo se realizan mediante la comparación de los resultados de las muestras extraídas, con los de muestras sembradas de una matriz similar y/o con los de blancos sembrados.

7.0 Análisis Técnico del Dominio de Aplicabilidad

Deben existir datos adecuados del método de prueba de sustancias y/o productos químicos representativos de aquellos que son relevantes para el CDU específico para el que se propone la prueba. La documentación de la MNE, ED o IATA deberá describir con claridad las propiedades fisicoquímicas del dominio de aplicabilidad. Esto incluiría cualquier limitación del método, por

ejemplo para sustancias químicas de un rango específico de peso molecular, volatilidad, solubilidad o estabilidad.

Si bien el dominio de aplicabilidad debe evaluarse analíticamente durante el desarrollo inicial de un ensayo, es posible disponer de datos adicionales mediante un uso más amplio del método. Estos datos podrían revelar interferencias que tal vez debieran abordarse más técnicamente (Petersen et al., 2022a). Si un ensayo no se evalúa para sustancias químicas con ciertas propiedades fisicoquímicas, habrá una mayor incertidumbre con respecto al desempeño del ensayo para estas sustancias químicas, tanto por cuestiones analíticas (por ejemplo, si existen sesgos que afectan el desempeño del ensayo) como de potencial concordancia (por ejemplo, si el ensayo produce resultados similares a un ensayo *in vivo*). Por lo tanto, también puede ser valioso evaluar la concordancia del ensayo entre sustancias químicas con diferentes propiedades fisicoquímicas, para aumentar la confianza en la aplicabilidad del ensayo con una variedad más amplia de sustancias químicas.

8.0 Control Positivo

Al menos un control positivo deberá ser parte del desarrollo del ensayo. Estos controles se pueden usar en conjunto con curvas estándar o en una sola concentración para el control positivo. Los controles positivos deben ser relevantes para el criterio de valoración de detección del ensayo. La concentración del control positivo debe estar dentro del intervalo de detección del ensayo. A menudo se elige una concentración de control positivo que produzca una respuesta entre el 50% y el 80% del rango máximo de detección, si se prueba una sola concentración y se dispone de datos sobre la resolución de la dosis máxima. Con frecuencia se produce una curva estándar utilizando el control positivo del sistema, definiendo curva estándar como la relación entre la respuesta del ensayo y concentraciones conocidas del analito. La curva estándar debe ser reproducible a lo largo del tiempo. Se deben preparar las curvas estándar en el mismo vehículo o solvente que las muestras de prueba previstas del estudio. Cuando no se puede obtener la misma mezcla o matriz, es posible emplear mezclas o matrices sustitutas, con la debida documentación y justificación.

9.0 Estándares de Referencia para Calibración de Instrumentos

Cuando corresponda, los instrumentos deben calibrarse utilizando estándares de calibración y/o muestras de control de calidad. Se debe documentar la fuente de la muestra estándar o de control de calidad, incluyendo pureza, estabilidad, fuente, número de lote, certificado de análisis y fecha de caducidad. Existen varios tipos de estándares, tales como los estándares compendiales de la Farmacopea de los Estados Unidos y materiales comerciales obtenidos de una fuente comercial de buena reputación con pureza documentada.

Idealmente, el estándar de control para algunos instrumentos (por ejemplo, espectrómetros de masas) debe ser idéntico al analito de interés en el ensayo. A menudo esto no es posible y se puede emplear una forma química establecida de pureza conocida.

10.0 Configuración de Especificaciones

Se pueden establecer especificaciones para medidas de control en proceso basadas en los resultados de pruebas intralaboratorio o interlaboratorios. Existen múltiples enfoques que se pueden utilizar. Por ejemplo, los métodos comúnmente empleados aplican un intervalo de confianza del 95% a las medidas de control en proceso o utilizan la media \pm dos o tres veces el valor de la desviación estándar, pero a menudo también se utilizan otros enfoques estadísticos. La configuración de especificaciones implica equilibrar diferentes objetivos. Es importante tener especificaciones suficientemente rigurosas para excluir datos que indiquen que el ensayo no está funcionando como se esperaba o que los datos de una medición de control en proceso están dentro de un rango que pudiera sesgar el resultado de la sustancia de prueba. También es importante no configurar especificaciones tan estrictas que las pruebas con datos en lo que se consideraría un rango normal, pudieran tener un número demasiado alto de ejecuciones del ensayo que no cumplan con las especificaciones.